DOCUMENT RESUME

ED 065 442 24 SO 004 683

AUTHOR	Heller, Jack J.; Campbell, Warren C.
TITLE	Computer Analysis of the Auditory Characteristics of Musical Performance. Final Report.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-9-0546
PUB DATE	May 72
GRANT	OEG-0-9-160546-4439-010
NOTE	131p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Analog Computers; *Auditory Perception; Aural Learning; *Computer Assisted Instruction; Digital Computers; *Listening Comprehension; *Music; Musical Instruments; *Music Education; Research; Research Methodology

ABSTRACT

The purpose of this research was to perform computer analysis and modification of complex musical tones and to develop models of perceptual and learning processes in music. Analysis of the physical attributes of sound (frequency, intensity, and harmonic content, versus time) provided necessary information about the musical parameters of intonation, vibrato, dynamics, and rhythm. A general purpose digital computer and appropriate analog devices were utilized to analyze and synthesize complex musical tones. The procedures included the transformation of audio tapes of music to digital tapes via a high speed analog-to-digital converter system. The significance of the research is based on the belief that: 1) objectifying certain parameters of musical performance will have a direct bearing on behavioral goals and methods of music education; and, 2) an understanding of the total problem of human information processing requires a detailed investigation of structured non-verbal stimuli in the auditory mode. Two supplementary investigations are included: Computer Analysis of Musical Performance, by Warren C. Campbell (Appendix I) and The Effects of the Attack Transient on Aural Recognition of Instrumental Timbres, by Ralph C. Thayer, Jr. (Appendix II). Appendix III, Computer Analysis System Software, by Jack Owens, describes a library of computer programs used to perform certain basic mathematical analyses of the digitized musical performances. A related document is ED 058 745. (Author/JMB)

ED 065442

BEQ-0546
PA 24
NCERD

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to: SO EM

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

ED 065442

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FINAL REPORT
Project No. 9-0546A
Grant No. OEG-0-9-160546-4439(010)

COMPUTER ANALYSIS OF THE AUDITORY CHARACTERISTICS OF MUSICAL PERFORMANCE

Jack J. Heller
Warren C. Campbell
The University of Connecticut
Storrs, Connecticut

MAY, 1972

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION AND WELFARE

Office of Education
Bureau of Research

CONTENTS

Page

LIST OF TABLES iv

LIST OF FIGURES v

SUMMARY vi

I. INTRODUCTION 1

Music Analysis 1
Development of Models 4

II. RESEARCH DIRECTIONS: EXPERIMENTS AND MODELS 7

A. Performance Adjudication and Analysis 7

Statement of Purpose 9
Statement of Problem 9
Method of Solving the Problem 9
Significance 10
Evaluation of Student Achievement 11
Application to Computer Assisted Instruction 12
Methodology 13
Psychometrics and Statistics 13
Pattern Recognition 14
Musical Acoustics and Special Equipment 14
Summary 16

B. Performance Modification and Listener Response 17

Introduction 17
Related Literature 18
Procedures 20
Results 20

C. Information Processing and Musical Perception 22

D. Systems for Performance Analysis and Modification 30

The Computer Analysis System 30
Tone Line Writer 32
Tone Line Reader 32
Harmonic Synthesizer 34

REFERENCES 36

APPENDIX I: COMPUTER ANALYSIS OF MUSICAL PERFORMANCE 41

Purpose and Problem 41

Procedures 41

Performance Selection and Preparation 42
Criterion Scores 44
Predictor Variables 45
Analysis 50
Summary 51

CONTENTS (continued)

APPENDIX I (continued) Page

Results 51

Analysis of the Criterion Variable 51
Analysis of the Predictor Variables 58
Multiple Regression Analysis 58
Summary 66

Discussion and Conclusions 68

Criterion Scores 68
Standards for Evaluating Computer Performance 68
Predictor Simulation of Criterion Scores 70
Generalizability 71
Future Research 71
Summary 72

References 74

Appendix A: Information Given to Judges 75

Appendix B: Data Normalization and Reduction 78

APPENDIX II: THE EFFECT OF THE ATTACK TRANSIENT ON AURAL RECOGNITION OF INSTRUMENTAL TIMBRES 80

Introduction 80
Problem 81
Limitations 81
Hypotheses 82
Procedures 82
Test 83
Scoring 85
Results 85
Summary of Results 93
Conclusions 94
References 96

APPENDIX III: COMPUTER ANALYSIS SYSTEM SOFTWARE 103

Programming Considerations 103
Implementation 104

APPENDIX IV: SCANNER/SYNTHESIZER INTERFACE 115

APPENDIX V: HARMONIC SYNTHESIZER 118

LIST OF TABLES

Page

APPENDIX I: COMPUTER ANALYSIS OF MUSICAL PERFORMANCE

Table 1. Characteristic Features, Calculated for Each Tone 46
Table 2. Predictors Used for Subset Calculations 50
Table 3. Interjudge Correlations 52
Table 4. Correlation of Each Judge With Average of Other Judges' Scores 55
Table 5. Summary of Inter-Judge Correlations, Inter-Group Correlations 55
Table 6. Regrade Correlations for Three Judges 57
Table 7. Intercorrelations of Performance Judging Categories 57
Table 8. Correlations of Predictors with Criterion Scores, Set #1 59
Table 9. Correlations of Predictors with Criterion Scores, Set #2 60
Table 10. Computer Simulation of Human Judgements 61
Table 11. Correlation of Simulated Scores with Criterion for Performance Subsets 63
Table 12. Cross-validation of Random Subsets 65
Table 13. Cross-validation of Natural Subsets 67
Table 14. Criterion Correlation Ranges 69
Table 15. Homogeneity and Reliability for Sixteen Judges 69

APPENDIX II: THE EFFECT OF THE ATTACK TRANSIENT ON AURAL RECOGNITION OF INSTRUMENTAL TIMBRES

Table 1. Means and Standard Deviations 86
Table 2. Means Converted to Percentages 36
Table 3. Results of Statistical Tests of Significance 87
Table 4. Statistical Tests Between Groups 89
Table 5. Correct Responses (In Percents) For Group C 89
Table 6. Response Errors for Group C 90
Table 7. Tables of Errors 97
Table 8. Tables of Percentages 100

LIST OF FIGURES

Page

II. RESEARCH DIRECTIONS: EXPERIMENTS AND MODELS

Figure 1. Performance Analysis 8
Figure 2. Learning Windows for Speech and Music 29
Figure 3. Brain Processing Models 29
Figure 4. Computer Analysis System 31
Figure 5. Tone Line Writer 33
Figure 6. Tone Line Reader 35

APPENDIX I: COMPUTER ANALYSIS OF MUSICAL PERFORMANCE

Figure 1. Overview of Procedures 43
Figure 2. Sample Chart Recorder Output, Soprano 49
Figure 3. Sample Chart Recorder Output, Alto 49

APPENDIX III: COMPUTER ANALYSIS SYSTEM SOFTWARE

Figure 1. Program Configuration and Logic for Use with the 2250 Display Console 102
Figure 2. Obtaining the Wavelength from a Resonance 106
Figure 3. Updating Disk Data Records 108

APPENDIX IV: SCANNER/SYNTHESIZER INTERFACE

Figure 1. Scanner/Synthesizer Interface (Schematic) 114
Figure 2. Scanner/Synthesizer Interface (Operational) 117

APPENDIX V: HARMONIC SYNTHESIZER

Figure 1. Schematic Diagram #1 119
Figure 2. Schematic Diagram #2 119
Figure 3. Schematic Diagram #3 119
Figure 4. Schematic Diagram #4 119
Figure 5. Data Samples for f_1 through f_{11} 123
Figure 6. Data Samples Before and After Filtering 124
Figure 7. Distortion at Breaks 124

SUMMARY

Investigations in speech have shown the complex nature of the processing which enables us to decode the acoustic signal of spoken language. Constraints on natural language are imposed by both the production and receiving mechanisms. The perception of musical sounds is obviously subject to some of the same constraints, since the same receptors and, possibly, some of the same processing mechanisms are used for music as for natural language.

The ability to identify and classify significant perceptual parameters of musical performance, and to state the effects of these parameters in behavioral terms is a prerequisite to the establishment of effective training procedures and meaningful behavioral goals for music education programs. Knowledge of these parameters will enable objective models of musical performance to be developed and verified.

The purpose of this research was to perform computer analysis and modification of complex musical tones and to develop models of perceptual and learning processes in music. Useful models of aural perception in music have been verified by comparing the responses of a computer implementation with the responses of appropriate human listeners. Analysis of the physical attributes of sound (frequency, intensity, and harmonic content, versus time) provided necessary information about the musical parameters of intonation, vibrato, dynamics, and rhythm. Information has been provided regarding the relative importance of attack and steady-state portions of tones for specific instruments. Further data provided information regarding brain hemisphere dominance for tasks of identifying instrument attacks when two different instruments are presented dichotically.

In order to analyze and synthesize complex musical tones a general purpose digital computer and appropriate analog devices were utilized. The procedures included the transformation of audio tapes of music to digital tapes (numerical data) via a high speed analog-to-digital converter system. Multivariate statistical techniques were used to provide improved psychometric capability in the perceptual domain.

The significance of this research is based on the belief that 1) objectifying certain parameters of musical performance will have a direct bearing on behavioral goals and methods in music education, and 2) an understanding of the total problem of human information processing requires a detailed investigation of structured non-verbal stimuli in the auditory mode.

I. Introduction

The purpose of this research was to perform computer oriented analysis and synthesis of complex musical tones, and to develop models of perceptual and learning processes in music. This effort is motivated by the belief that the ability to identify and classify significant perceptual parameters of musical performance, and to state the effects of these parameters in behavioral terms is a prerequisite to the establishment of effective training procedures and pertinent behavioral goals for public school and college music programs. Knowledge of these parameters should enable objective models of musical performance to be developed and verified. This is, of course, not a new idea. One of the earliest proponents of this point of view is Carl E. Seashore (Psychology of Music, McGraw Hill, 1938). Additional impetus is given to this work by related experimentation in linguistics and psychology which point out that an understanding of the total problem of human information processing requires a detailed investigation of structured non-verbal stimuli in the auditory mode.

Music Analysis

The auditory characteristics of musical performance can be objectively examined in terms of the physical characteristics of the complex acoustic wave generated by the performer. For this research, the fact that there are visual aspects and other contingencies that may influence listener response to a performance is specifically not considered. This restricted situation is comparable to an individual listening to a tape or disc recording: the sound pattern can be completely represented at any of several points in the auditory channel as the variation of a single parameter plotted against time. Traditionally, if the vibration is reasonably repetitive, this information is reduced by introducing the concepts of frequency, intensity, harmonic content, and spectral distribution of noise, as functions of time. Their use allows for significant data reductions without serious loss of information of importance in the analysis of music.

Every perceptually significant nuance of the auditory component of musical performance, no matter how complex, can be described as a function of these attributes of sound waves. This statement should not be read as minimizing the extreme complexity of the subtleties and contextual dependencies involved in artistic performance.

Many laymen and musicians believe that attempts to analyze artistic performance are futile. They contend that individual difference is an overriding characteristic of artistic performance. The importance of individual difference is not denied, particularly when considered as the hallmark that permits performer identification. However, there is a need to understand the constituent elements and limits of variability common to all artistic performance, as a prerequisite to any discussion of the effect on the listener of individual difference.

Musicians probably will agree, that before one can produce an artistic performance, there are certain technical demands of performance which must be satisfied. Control of these technical aspects, such as intonation, dynamics, tone quality, rhythm, and attack, is a prerequisite to performance at a professional level.

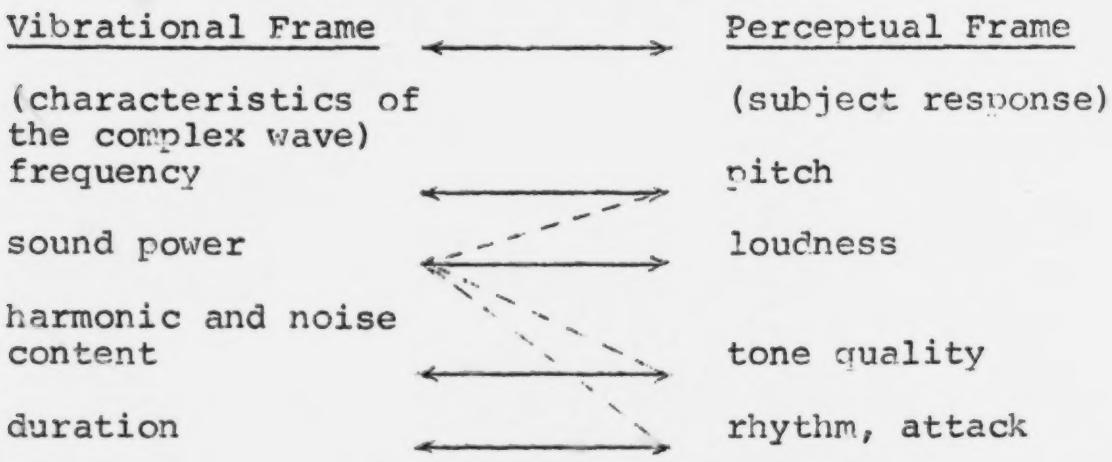
An analogy can be drawn between the musician and the literary writer. The latter cannot create a literary masterpiece until he has learned to control language. The writer who achieves optimal creative production through written communication must first master the language and make it subject to his control and manipulation. It is the same in the creative production of music. The artists who achieve a high level of performance in music will emerge from that group of musicians who have refined the technical aspects of performance.

This project was in no way an attempt to mechanize the creative process or to place restraints upon those persons who are capable of making unique contributions to music. Rather, the attempt was to identify the perceptually significant aspects of music by analyzing model performances that exemplify a high level of musicianship (as judged by professionals), and by comparing these to performances that fall short of this level.

The basic analysis may be thought of as a

mapping of auditory characteristics from the perceptual to the vibrational frames of reference, and the reverse. For each perceptual event, a vibrational counterpart or correlate is sought which can be described in terms of the parameters indicated below:

AUDITORY MAPPING



The relationships indicated by the solid lines between the vibrational and perceptual parameters are first-order effects. Functional relationships and difference limens were established for some of these main effects by the early experimenters in audiology. For example, the relationship between sound power or intensity and loudness can be roughly represented by converting power level readings in watts to the decibel scale:

$$\text{loudness (db)} = 10 \log_{10} \frac{P}{P_o},$$

where P is the sound power, and P_o is a reference level. An approximation to musical pitch can be obtained by converting frequency to a measure representing pitch in the tempered scale:

$$\text{tempered pitch (semitones)} = 12 \log_2 \frac{f}{f_o}$$

where f_o is a reference pitch of zero, and the octave above f_o ($f = 2f_o$) is represented as "12" on the semitone scale. The same procedures can be followed for approximating musical pitch to other scales (Pythagorean, Just, etc.).

These transformations are attempts to represent the perceptual domain in terms of mappings from the vibrational domain. They are "first-order" approximations because they take into account the most important perceptual - vibrational links. However, the nuances of musical performance require that second- and perhaps third-order effects be represented in the mapping if prediction of the musically significant responses to a performance are to be achieved.

A known second-order effect is the change in pitch introduced by an intensity change in a tone held at constant frequency. Second order effects can be categorized as the predictable changes in one perceptual dimension due to changes in the vibrational correlate of another perceptual dimension. Seashore calls them the "normal illusions" of auditory perception. In terms of the diagram of auditory mapping, these effects would be indicated by the dashed lines connecting different levels (only those leading from "sound power" are shown). While many of these effects have been investigated for pure tones and static situations, their importance and application in a musical context have not, in most cases, been delineated.

Development of Models

Investigations in speech have shown the complex nature of the processing which enables us to decode the acoustic signal of spoken language. Constraints on natural language are imposed by both the production and receiving mechanisms. The perception of musical sounds is obviously subject to some of the same constraints, since the same receptors and, possibly, some of the same processing mechanisms are used for music as for natural language.

Perceptual parameters such as pitch, duration and tone quality, that are a regular part of the musician's vocabulary, are basic to an understanding of language processes (Lieberman, 1967). While the ear is the common receptor for speech, music and noise, the possibility has been raised that more than one mode of processing may be operating, that a speech/non-speech dichotomy may exist. The utility of basic brain-function

models can be better evaluated when data from research in music/speech perception are presented. For example, "hemisphere dominance" appears to be a uniquely human characteristic, which is a vital part of the human linguistic capability. One of the features associated with this phenomenon is the differential processing of auditory signals, depending on the type of signal input.

Music of all types consists of highly organized sets of complex auditory relationships which are processed by humans. An understanding of the feature extraction and organizing function of the brain for non-verbal auditory phenomena such as music may provide clues to the building and refining of models of human information processing. The pattern recognition and processing capability in language has been studied extensively. However, studies in linguistics and psychology have not developed testable models of information processing which isolate the unique characteristics of music.

While the evidence suggests that the auditory processing mode is dependent upon the type of input, the status of music in this apparent dichotomy has not been established. If there are two modes for processing auditory signals, the possibility exists that a perceptual disability, such as autism, may selectively interfere with only one of these processing modes. Knowledge of this condition may allow for the development of early diagnostic tests for the severity of the disability, and at the same time indicate the possibility of new communication techniques to improve the compensatory use of intact processing modes.

Knowledge of auditory processing modes is basic to an understanding of the acquisition of linguistic and musical skills by the normal child. The acquisition of reading skills and the development of tonal memory are basic educational problems certain to be affected by a knowledge of processing modes. In the course of normal development, the auditory processing mechanism apparently undergoes important changes up to about the age of five (Kimura, 1967). Beyond this age, at least with respect to auditory pattern recognition, we may all be "perceptually handicapped". The learning of basic sounds of language must occur before a

critical age for the discrimination to be fully integrated into the language function. On the basis of results reported by Kimura (1964), it is expected that this will also be true for music. Detailing the differences between the perception of speech and non-speech has begun to lead to specific models for the understanding of this developmental change.

II. RESEARCH DIRECTIONS: EXPERIMENTS AND MODELS

A. Performance Adjudication and Analysis

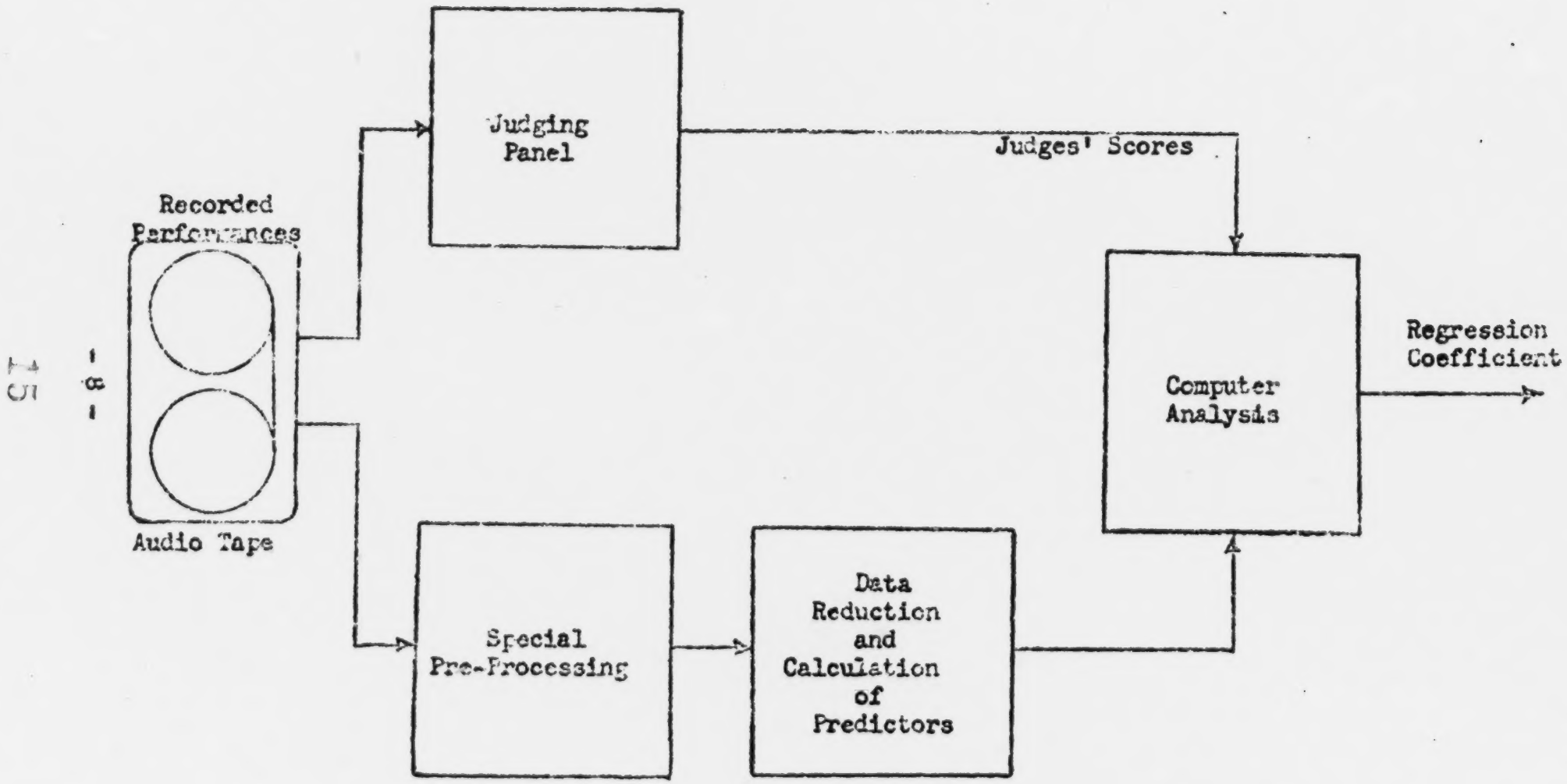
Evaluation of student achievement is a central problem at all levels of Education. Grading student output is a major chore for most teachers, and is often the one in which they find the least satisfaction. The problem is a particularly difficult one in music education, where the musical performance is the focus for student achievement. Experienced music teachers and performance judges contacted in the course of this project were skeptical of their own ability to maintain objectivity and uniformity under many of the conditions encountered in practical adjudication situations. No studies were found which dealt directly with the problem of the reliability of this type of evaluation of musical performance. Information is required concerning both the regrade reliability of a single judge, and the inter-judge agreement for groups of judges when dealing with various types of musical performances.

Studies of this type have a long history in the area of the grading of student essays. For example, studies by Findlayson (1951) and Phillips (1948) show a self-correlation among graders of between .60 and .70. That is, when a teacher is asked to regrade a set of essays after a time-lapse, the agreement between the two sets of grades is only 36 to 49 percent better than would be expected on the basis of chance alone. Interjudge correlations are in general even lower than this. Page and Paulus (1968) report inter-judge correlations ranging from .43 to .59 obtained by comparing the grades given by experienced judges on a set of student essays. An effort to improve the reliability of essay grading in an effective and practical way was reported by Page (1966) in an article entitled "The Imminence of Grading Essays by Computer."

The approach taken by Page uses a simulation of the human grading process through the use of a digital computer. A computer analysis of essay features that can be given an actuarial representation is used to define the score for a given essay. The averaged scores from several human judges is used as a criterion, since it has a higher

Figure 1.

PERFORMANCE ANALYSIS



reliability than the score given by a single judge, when the judges are all from the population of interest. The effectiveness of this approach, reported by Page and Paulus (1968), provided the stimulus for a similar application to the grading of musical performances, which is the subject of this section of the report. A pivotal premise for both studies is the possibility of defining variables which are useful correlates for the subjective variables influencing the human judge as he rates a given example of student output.

Statement of Purpose

This investigation is an attempt to add to the body of knowledge related to the following questions:

1. Are there operationally definable (i.e., objective, measureable variables for which musical performance limits can be established?

2. If so, what are they and where are the limits of acceptability for various performance situations?

3. Can a knowledge of such variables be made useful in solving the problems of music education?

Statement of Problem

Given a set of scores assigned by a group of competent human judges to a set of student musical performances, are there objective features of the recorded performance which can be used, with suitable computer analysis, to predict the averaged judges' score?

Method of Solving the Problem

The following steps were taken in this investigation: (See Figure 1.)

1. A sample of 62 recorded performances was selected from the 1967 Connecticut All-State music auditions.

2. A group of seven competent judges was hired to grade each performance on five categories. The

judges' averaged score in each category for each performance served as the criterion measure for computer grading.

3. Six objective features of the performances thought to be potentially useful for predicting performance quality were selected on an a priori basis. Specialized machine and hand techniques were developed to process the performances to extract the features. Values of the features associated with each note of each performance were punched on computer cards.

4. The digital computer was employed to process the feature values for normalization, recovery of initial data, and data reduction. Thirteen predictor values for each performance were derived from the original six features.

5. Two different performance standards were used in deriving the predictors, and the results compared.

6. The thirteen predictor variables were entered into a multiple regression program on the digital computer, and weightings were assigned to each variable in order to maximize the correlation between the predicted and the criterion grade.

7. The correlations were empirically cross-validated on subsets of the original sample in order to determine how well this program could be expected to predict scores for a new sample of performances.

Significance

The investigation of objective features of performance to determine their usefulness in predicting judges' responses is of importance in at least two major areas of music education, (i) the evaluation of student achievement and (ii) the development of practice and diagnostic sessions in which computer generated feedback supplements regular tutorial procedures (i.e., Computer Assisted Instruction).

Evaluation of Student Achievement

The evaluation of student musical performance by human judges would appear, from the results of this investigation, to be at least as reliable as the evaluation of student prose. Further confirmation of this result is required. The computer, given a valid regression equation, will produce an identical evaluation each time a particular performance is submitted and will apply ratings without the biases introduced into human judging by fatigue, effects of performance juxtaposition, knowledge of student personality, sex, race and appearance.

A successful simulation of the judging of musical performances by human auditioners has an immediate application to the problem of auditioning large numbers of student musicians for placement in schools and musical programs. Many states have "All-State" band, chorus and orchestra festivals to give experience and encouragement to music students. The process of auditioning hundreds of these students and to give each a fair chance to demonstrate his ability is a very difficult one. Computer grading may well provide the answer to this problem.

A direct comparison of student achievement or evaluation techniques between groups separated in either place or time has not been practical in the past. For example, it might be desirable to know how this year's students would be rated by last year's judges. The judges, as a group, may no longer be available. However, if an adequate sample of their scoring is available, a valid computer simulation would provide an accurate representation of their collective opinion on this year's performances, or on any other set of recorded performances.

This procedure is not limited to a particular standard of judgement. With judges representing several schools of thought or stylistic preference, categorization of performers with regard to stylistic suitability could be done effectively and efficiently. For example, a choral conductor might select an ideal grouping for a main and an echo choir, or, choose the voices most appropriate for a madrigal group, without the necessity

of individual auditions each time a new partitioning of the chorus is required.

Application to Computer Assisted Instruction

Kuhn and Allvin (1967), reporting on work by Spohn (1959) and Carlsen (1963) in the area of programmed instruction state (p. 2):

> "As these two examples indicate, programing techniques for music instruction have been successfully undertaken as long as essentially verbal or symbolic response modes have been used, such as multiple choice, true-false, comparison, notation, and so forth. Typical response modes in such studies are marking, pointing, using a typewriter, and so forth.
>
> Obviously, a new, direct, and essentially musical response mode is needed."

These two authors go on to describe an experiment in CAI in which a musical response mode is an essential part. Student vocal response was analysed with a "pitch extractor", and the resulting information was fed into an IBM 1620 computer for processing. Evaluation was made in terms of deviation limits from the true pitch, using equal temperament. Students were able to select limits of either a two percent or four percent range.

Deihl and Radocy (1969) report a similar effort, using an IBM 1500 Instructional System. Here, however, student response was off-line, but mention is made of the possibility of future on-line interaction.

These examples indicate the directions being taken in the application of CAI to music education. However, none of these studies comes to grips with a fundamental problem common to all attempts at computer interaction with the performer: the relationship between the acoustic variables being measured, and the subjective responses of competent listeners. Training in the control of an acoustic variable does not necessarily result in the musical control of a subjective variable.

The approach used in this investigation provides a test of the relationship between the acoustic parameters chosen, and the subjective responses indicated by the judges. Once a stable relationship has been discovered, a training procedure based on manipulation of the acoustic parameter can be developed, and the results can be anticipated with some assurance of success.

Methodology

The simulation techniques used in this experiment were patterned after those used in Project Essay Grade (Page and Paulus, 1968), an investigation supported by the U.S. Department of Health, Education and Welfare (Project No. 6-1318). Differences which occur between this project and "Project Essay Grade" are those imposed by the different mediums being explored, rather than any difference in approach. These appear primarily in the area of feature extraction. For the essay grading project, the data, consisting of handwritten student prose, was punched as literally as possible into IBM cards. Features were extracted by a computer analysis of the text: spelling, punctuation, word length, sentence length, etc. In the music project, features are extracted from a recording on magnetic tape of the performance to be analysed. For the essay grading project, judgements in five categories were predicted for 256 essays using thirty predictors. In the music project, judgements in five categories were predicted for 62 vocal performances using thirteen predictors.

Psychometrics and Statistics

The basic conceptual framework for evaluation and prediction in this investigation is found within the disciplines of psychometrics and statistics, as presented in many fundamental texts such as those by Hays (1963), Kelley (1948) and Winer (1963). Rozeboom (1966) was particularly useful in his discussion of homogeneity, reliability and validity. The programs for correlation were based on Cooley and Lohnes (1962). The stepwise program from the System/360 scientific subroutine package (IBM, 1968) was used for the multiple regression analysis.

Pattern Recognition

In a general sense, a simulation which involves responses to an input data set is a problem in the field of pattern recognition. In this case a pattern must be detected in the acoustical signal which can be used to predict the response of musical judges to that signal. A presentation of basic concepts in pattern recognition may be found in Sebestyen (1962) and in Nilsson (1965). Statistical models are an important part of pattern recognition procedures, with greater emphasis on decision making than is usually found in statistical usage. The extraction of features from which decision boundaries can be constructed is another fundamental process in pattern recognition, but there are no general algorithms for extracting useful features from a data base. Feature definition must be based on the attempts of previous investigations to find measurable aspects of the data field. Features which are useful for prediction are often specific to a given application. To be useful, features which are based on a nominal or ordinal scale must correlate, either separately or in combination, with the criterion. Nominal features must be evaluated using categorical decision processes. For this investigation, the features extracted were all given interval scale interpretations, and were based in part on the findings of studies in musical acoustics and the psychology of music, discussed in the next section.

Musical Acoustics and Special Equipment

The development of techniques for operationally defining important acoustic variables owes its greatest advancement in the 19th century to H.L.F. Helmholtz (1877). While others in this and earlier centuries laid the foundations for vibrational phenomena (Young (1784), Rayliegh (1877), Fourier (1822), etc.), Helmholtz explored the perceptual problems as well. Of the many recent publications in musical acoustics, books by Benade (1966) and C.A. Taylor (1965) have been particularly helpful in the present investigation.

At the turn of the century, the career of

Carl E. Seashore, who dominated the field until World War II, was just beginning. A frequently stated objection to many experiments in the psychology of music is that they are conducted in such a way as to remove them from the reality of musical performance. Studies in perception using isolated tones, pure tones and other convenient simplifications do not always generalize to real performance situations. Seashore tried to avoid this problem by working primarily from "live" as opposed to laboratory situations. The same attempt at relevance is made in this project.

In one of his first papers, A Voice Tonoscope (1902); Seashore described a device that was able to present visually, in graph form, the fundamental frequency of a live performance as a function of time. This device made it possible to measure systematically, variables thought to be related to musical perception, and to test hypotheses about them.

The equipment used to process the acoustic signal in this investigation, a modern version of the tonoscope, was planned and tested by J. Heller and W. Campbell. Its use in a related application, visual pitch matching, is described in a report by J. Heller (1969). A standard Frequency Modulation Sub-carrier Discriminator (Electro-Mechanical Research #287A-01), normally used in telemetry applications, was used as a frequency meter. It was modified by the manufacturer to operate over a pitch range of one octave from slightly above middle C to the next octave (266.7 Hz to 533.3 Hz, or 400 Hz $\pm$ 33%). Using a Schmidt trigger, the discriminator produces a DC output proportional to the input frequency, over the input range: $v = k(f - f')$ where v = output voltage, f = frequency of input signal, f' = center frequency, in this case 400 Hz, and k is the proportionality constant. The output voltage was presented as a function of time, on paper charts using an Esterline angus Speed Servo Recorder/S601 S.

A Sony Model TC-5600 Tape Recorder, with variable speed control, was used to adjust the performance pitch range to the acceptance range of the frequency meter. By noting the actual starting pitch for each performance, and entering these data into the computer program, all pitch and

time values were restored to correspond to the original performance speed. These calculations are detailed in Appendix B.

The choice of features which are related, at least indirectly, to the perceptually significant aspects of the acoustic signal is an essential part of this approach. Clues to the nature of these features can be found in the many sources in the psychology of music. Seashore (1938) provides a detailed account of the relationships between the objective and perceptual variables in his chapters on Pitch, Loudness, Duration and Timbre. He also enumerates (Seashore, 1938, p. 28) a list of basic principles in the psychology of music which are taken, with only minor changes, as fundamental concepts for this investigation. More recent studies (J.D. Harris, (1952), J.C.R. Licklider (1956), Robinson and Dadson (1956)) have provided details and further clarification of the illusions of hearing referred to by Seashore.

Summary

The first experiment completed under the grant was a computer simulation of an adjudication of musical performance. In this investigation, a panel of seven competent musical judges was asked to audition a set of sixty-two short vocal performances by students, which were recorded on audio tape. The judges were to respond by scoring each performance on a five point scale in several categories, such as intonation, dynamics, etc. The average of the judges' scores was found to be stable, and was the criterion for a computer simulation of the judges' response, using multiple regression analysis. The prediction of the judges' response to each performance was based on frequency, intensity and duration measurements of the performance.

This approach provides a means of examining the relationship between performance competence, as rated by experienced judges, and the vibrational characteristics of performance. Specific vibrational patterns can be tentatively identified as correlates of a particular response from the panel of judges. Synthesis of these vibrational patterns can then be used to test the possibility of a cause and effect relationship between the

pattern and the response. The results of this experiment were very encouraging, particularly in view of the highly simplified (linear) predictive model employed.

Some idea of the simulation capability achieved in this first attempt can be seen in the following table. The numbers indicated are typical correlation values. A correlation of unity means perfect agreement, and a value of zero indicates a random relationship.

Score Comparison	Correlation
Between two groups of judges (group averages)	.85
Between computer scores & judges' average	.65 (shrunken mult-r)
Between one judge and the group average	.65
Between two judges	.40

From this it can be seen that the computer comes as close to predicting the group average as the typical experienced judge (.65). The correlation between any two judges (.40) is considerably lower than this value. Refinements in technique are expected to further improve computer prediction of judges' scores. A complete report of this experiment appears in Appendix I.

B. Performance Modification and Listener Response

Introduction

A second experiment completed during the grant period was designed to determine the effect of the attack transient upon the recognition of instrumental timbres. This study approached the problem by mechanically replacing the attack of one instrument by the attack of another, to determine if the listener is influenced in his attempt to recognize instruments more by the attack or by the steady-state portion of the tone. This controlled modification of natural sounds was used to test hypotheses regarding differences between instrumental

timbre recognition for normal, altered, and "no-attack" tones.

Related Literature

Several previous studies of timbre have recognized the importance of the attack transient in timbre recognition. Seashore (1938) alludes to a study by Lewis and Cowan (1937) who mechanically replaced even vocal releases (decay) with gliding attacks. "The musically acceptable glide at the beginning of the tone became utterly intolerable when placed at the end of a tone." Seashore does not investigate the effects of attack or decay on sonance (tone quality), but states, "This experiment opens a very fertile field for the investigation of reasons for adaptive or habitual hearing."

Nolle and Boner (1941) found, in investigating the initial transients of organ pipes, that "these initial transients seem to be important in determining the subjective character of pipe organ music and in differentiating between the subjective character of pipe organ music and of music produced by electronic instruments."

In synthesizing instrumental tones, studies by Fletcher, Blackham, and Stratton (1962), Strong and Clark (1967), and Risset and Mathews (1969) have stated that the attack plays an important role in recognition of the instrument.

On the basis of analyses of the harmonic content of various wind instruments, with one result being that little substantial difference was found to exist between the harmonic content of certain instruments, Saunders (1946) concluded, "These facts lend support to the idea which is often expressed that an oboe and a violin would be indistinguishable if one were prevented from hearing the beginnings or the endings of the tones."

Richardson (1954), in an investigation of transients by spectrum analysis, states, "In spite of their evanescent nature, the view is now held that it is these transients which enable the listener to distinguish the sounds of different musical instruments or between two of the same class. The transient is indeed part of the 'formant'

of an instrument, and ought to be exhibited as a characteristic alongside the steady-state spectrum."

These studies have all indicated, to one degree or another, that some importance may be attached to the attack as an indicator of timbre. However, all of these studies were performed under conditions in which the evaluators were aware of what timbre was being considered. Under such circumstances an evaluator's judgment would tend to be based on an ideal concept of the particular tone being judged, rather than a mere identification of timbre.

A more valid approach to measuring the importance of attack to timbre recognition would be to present subjects with unidentified tones, with and without attacks, to be identified by the subjects. This method has been utilized in studies by Berger (1964), and Saldanha and Corso (1964).

Berger presented 30 subjects with the tones of 10 different wind instruments. The tones were presented in several forms: unaltered, played backwards, attack and decay removed, and all harmonics except the fundamental filtered out. The results showed 59 per cent of the unaltered tones identified correctly and 35 per cent of the tones minus attack and decay identified correctly. The tones played backwards and the filtered tones were correctly identified 42 per cent and 18 per cent of the time, respectively.

Saldanha and Corso, in a similar study, presented 20 subjects with the tones of 10 string and wind instruments, unaltered, and with various alterations. The results showed 41 to 44 per cent of the unaltered tones correctly identified (the two figures are a result of longer and shorter steady-states), and 32 per cent of the tones with no attack identified correctly. Both of these studies indicated a strong relation between attack and timbre identification, but no statistical analyses of the results are provided to confirm this fact, nor is the effect of the quality of the attack investigated.

In spite of the work which has been done in this field, one finds that the generally accepted definition of timbre remains as some exclusive

function of the harmonic content. Lundin (1967) discusses timbre strictly in terms of the steady-state portion of the tone, and Neilson (1970) describes timbre as the "relationship of the decibel strength in the fundamental of a given tone to that of various overtones."

The present study attempts to demonstrate whether or not the initial transient is an integral part of timbre recognition, and, if it is, to investigate what effects changes in the quality of the attack have on the recognition of timbre.

Procedures

A stimulus tape was prepared which included 120 tones. Four instruments, (flute, oboe, clarinet, and trumpet) each playing three pitches, (d',c'', gb'') were recorded. Each of these 12 tones was modified by replacing the normal attack by the attack of the other three instruments on the same pitch. A second modification of these 12 initial tones was made by eliminating the attack portion of each tone. This produced 48 tones. The original 12 tones (with no modification) was also included in the final tape. Each of these 60 tones were recorded a second time and placed on the final tape in a random sequence.

The stimulus tape was administered to three groups of subjects, high school instrumentalists (n=57), college students enrolled in an introduction to music history course (non-music majors, n=43), and college music majors including several music faculty and professional musicians (n=38).

Results

The flute and clarinet steady-state tones were identified correctly (82% and 79% respectively) more than the oboe and trumpet steady-state tones (70% and 77%). The identification of the flute was slightly less accurate than that of the clarinet when preceded by other attacks, but was more accurate when preceded by its own or no attack. The attack portions of flute and clarinet, however, are identified quite differently

from each other. The flute attack was least often correctly identified and the clarinet attack most often correctly identified. This indicates that the flute provides very strong identification information in its steady-state, but very little in its attack.

The oboe was the least recognizable instrument when presented without modification (70% correct). The trumpet was recognized somewhat better than the oboe (77% correct). With the removal of the attack, the trumpet dropped well below the oboe in degree of recognizability. With the addition of attacks of the other instruments oboe identification scores dropped considerably (from 63% to 51%), the trumpet scores only slightly (from 45% to 42%), nevertheless, the trumpet remained the least accurately identified tone quality. The oboe steady-state was most influenced by the attacks of other instruments, while the oboe attack had the least influence on the steady-states of other instruments.

The trumpet attack provided a great deal of identification information in combination with the trumpet steady-state. While in combination with other steady-states, the trumpet attack did not provide as much identification information as the clarinet attack, it provided more information than flute and oboe attacks. The trumpet steady-state by itself or in combination with other attacks was easily confused or identified as the attacking instrument, and the trumpet attack with other steady-states did not provide as much information as one might anticipate, in light of its effect on the trumpet steady-state.

The overall results of this study show that identification of timbre becomes less accurate (statistically significant beyond the .01 level) as the tone progresses from normal, to no-attack, to altered. That is, the attack affects aural recognition of timbre. (See Appendix II for a detailed discussion, and for tables of results.)

C. Information Processing and Musical Perception

It is a commonly held notion that speech and music occupy different levels of brain function. Speech perception is considered to be a highly ordered process, involving complex decoding and pattern recognition techniques. Music, by contrast is often considered to be a visceral activity, involving the lower brain centers, the emotions, and the rhythms of body movement and heart-beat.

This research is an initial attempt to test an entirely different model, in which music and speech are considered as analogous functions. That is, that musical perception is also a highly ordered process, subject to some of the same constraints and peculiarities that are a basic part of speech perception.

Man is the only mammal whose brain shows evidence of a strong functional asymmetry between the cortical hemispheres (Sperry, 1964). This unique adaptation, which reduces the redundancy and increases the functional capability of the brain, is closely related to the development of language. There is very strong evidence that between the ages of 1 and 6 in the human child, the language function is (normally) taken over by the left hemisphere, and that some musical and higher order visual processes are performed in the right hemisphere.

Evidence for the functional asymmetry of the human cerebral hemispheres comes primarily from four separate areas of investigation:

1. Tests of functional limitations in brain damaged patients for whom the location and extent of damage is known (Luria, 1970; Mountcastle, 1962).
2. Electro-encephalagraphic studies (Cohn, 1971).
3. Tests of epileptic patients who have undergone a sectioning of the corpus callosum (the inter-connecting structure between the cerebral hemispheres). (Gazzaniga, 1967).
4. Experiments on normal subjects using the technique of "dichotic presentation" (presentation of simultaneous but different stimuli to each

ear). The success of this technique in determining the locus of processing for a given input stimulus, is based upon the hypothesis that the representation for the stimulus presented to each ear is greater for the contralateral than for the ipsilateral hemisphere. (Shankweiler, Studdert-Kennedy, 1967; Kimura, 1961; Knox and Kimura, 1970). Many experiments with speech stimuli have been conducted using the dichotic technique, but only two have involved musical stimuli (Kimura, 1964; Shankweiler, 1966) presenting different melodies to each ear, with a left ear recall preference.

Because of the very limited nature of the investigations specific to music, and the many statements linking music variously with speech and non-speech (Warren, 1971) a pilot study was instituted in order to establish a basis for tentative brain function models in music. Shankweiler and Studdert-Kennedy (1967) showed a right ear superiority for consonants, but none for vowels. They stated "in view of Kimura's finding (1964) of a left ear advantage for musical melody recognition, as against a right ear advantage for spoken digits, the neutral status of steady state vowels, midway, as it were, between speech and music, is perhaps not surprising." (p. 60, Shankweiler and Studdert-Kennedy, 1967).

Melody recognition is of course only one of a large number of responses available to a moderately "literate" listener. It is possible, using plausibility arguments, to construct a comprehensive multi-level analogy between speech and music, from the phoneme level to the breath group and sentence. The analogy is,however, only an interesting exercise if it is not found to be in some way predictive of similar perceptual processes. The corresponding structural elements in music then, by implication, involve cues for decoding and defining the structure generating a particular musical "message".

The phoneme is the building block for natural language, with classification categories of vowels, consonants, semi-vowels, liquids, etc. In speech, the consonant has been shown to require a complex decoding process involving primarily the left

hemisphere (Broadbent and Gregory, 1964). Spectral envelope changes occurring within 20 to 60 milliseconds provide the primary identification cues (Liberman, et al., 1967). The musical analogy to the consonant/vowel/consonant sequence is the attack/steady-state/decay sequence for an instrumental tone. In speech, the consonant environment conditions the perception of vowels, and consonants provide a much greater reduction of uncertainty than do vowel sounds. A number of studies have shown that the attack portion of an instrument tone also has these attributes (Stumpf, 1926; Tenney, 1965; Risset and Mathews, 1969; Strong and Clark, 1967a, 1967b).

Since the phoneme is the fundamental speech segment, features such as length, stress and tone that span more than one segment are designated "suprasegmental". These suprasegmental features have direct counterparts in the musical phrase: stress, rubato, portamento, articulation, dynamic contour, timbre modification.

The musical counterpart to syntax has been extensively documented under the general category of melodic and harmonic structure.

In order to test the analogy it seemed appropriate to start at the most basic level, that of the phoneme. The method of dichotic presentation provides a clear operational test of the music/speech analogy. Are instrument attacks more readily identified when presented to the right ear, with the implication of left hemisphere processing?

To answer this question, six musical instruments (trumpet, violin, guitar, oboe, clarinet, flute) were recorded while producing a concert "middle-C" (nominal frequency, 261.6 Hz). These instruments were chosen to represent the major classes of sound production in music, with the exception of percussion. The violin was bowed and the guitar plucked. The flute was replaced with a small bottle tuned to 261.6 Hz, because of the very long attack transient for this low note in the flute's range.

The "attack" portion of each tone was defined by the experimenter on the basis of oscilloscope traces. The durations chosen were:

trumpet, 45 ms; violin, 43 ms; guitar, 33 ms; oboe, 47 ms; clarinet, 55 ms; "flute", 65 ms. The tones were recorded a second time with the steady-state portion of the original instrument replaced by an electronically produced triangular wave at approximately the same pitch and phase (at the transition point) as the original signal. Channel switching was accomplished using voltage controlled amplifiers, envelope generators and followers, trigger delays and oscillators from a Moog synthesizer.

The six hybrid sounds were then aligned and recorded on two separate tape tracks, so that all 15 pairings of dissimilar instruments were represented. A random sequence of 30 items (each stimulus presented twice, with channels reversed on the second presentation) was prepared, along with a single channel training sequence. Four subjects were taught to label the sounds reliably (the composite tones could be discriminated, but were not initially identifiable). They then listened twice to the random sequence of 30 stimulus pairs presented dichotically. For the second presentation, the earphones were reversed, to correct for any imbalance between the two electronic channels. Subjects were asked to identify both instruments in a stimulus pair, guessing if necessary. The scoring was based on responses for which only one instrument was correct. All subjects responded at well above the chance level: the average number of presentations (out of sixty) for which at least one instrument was correct was 52.

All subjects showed a right-ear advantage for identifying instrument attacks. The right ear average was 24.25 correct responses, compared to a 14.25 average for the left ear (difference significant at $p < .05$). This result lays the foundation for further experiments to detail the processing analogy between speech and music.

The research in speech over the last 20 years has progressed steadily on the problem of human information processing. The study of the perception of music has received little attention from this point of view, and comparative music/speech studies are difficult to find (Slawson, 1968; Kimura, 1964). This neglect is

possibly due to the caricature of music, fostered even by musicians as a "visceral", "non-intellectual" activity.

It is our contention that music is as complex a pattern recognition activity, even for the casual listener, as is speech. In fact, it seems reasonable to suppose that the structure of music may provide a more useful map of the brain's symbol manipulation process than speech, since fewer constraints (cultural, articulatory, referential) are imposed upon the development of musical patterns and modes of production. Implicit in this surmise is the idea that both speech and music map to subsets of a generating structure which is more general than the structures delimited by "speech" and "music". In addition, this general structure, while very large in terms of storage, and very flexible in terms of pattern organization, cannot be considered infinite. In 1963, H. Bremermann showed that the fundamental coarseness of matter does not allow it to transmit more than 1.6×10^{47} bits per gram per second. Ashby has pointed out that, for combinatorial interactions, even a very moderate-appearing situation, such as a square screen of 20 by 20 lamps, provides enormously more possible patterns than those that could be processed by any device the size of the brain, over a lifetime. The number of discriminable patterns or pattern classifications must be very small indeed, compared to these fundamental limits.

In speech the recognizable articulatory patterns are limited to about 60 phonemes for all the world's languages and only a subset of these is used in any one language. In addition, their identification is not fixed, but is, to some extent, context dependent. In music, the "phonemic" elements are tones of distinct pitch, timbre, loudness and duration. Sequences and combinations of listener"enjoyment" or appreciation is based on the ability to recognize the sequence of tones as belonging to a class of patterns previously delimited (in memory) by implicit rules of musical grammar. It should be emphasized that for the general listener, this recognition may not be recoverable; that is, he cannot specify the basis upon which his decision is made.

This is analogous to the language auditor who responds "Da" when "Ba" has been presented, but has no awareness of the distinctive features which he uses to discriminate "Ba" from "Da".

Referring to a higher linguistic level, Langacker (1968, p. 234) states:

> When the child has learned to talk, when he has mastered his native tongue, he is in possession of an abstract system of rules that specify an unbounded class of well-formed sentences. He is not conscious that he possesses this system in the sense that its structural patterns have been imposed on his psychological processes, so that these patterns are a factor in determining the course of his verbal activity. Learning to talk, like learning to ride a bicycle, involves the mastery of a set of principles; it involves the addition of structure to the body of psychological skill or competence that shapes our mentally directed behavior. These rules are thus no more accessible to conscious inspection than the rules for keeping one's balance while riding a bike. We talk and we keep our balance on bicycles, but in neither case do we know, at the level of consciousness, precisely what the guiding principles are.

It is clear that for language the developmental period from birth to about six years old is a very critical one, in the sense that sound patterns not learned during this period are only learned with great difficulty, if at all, at a later age. A similar argument applies to syntax, although at this level, usable ex post facto rules can be devised to approximate the structure "discovered" by every normal child before the age of five.

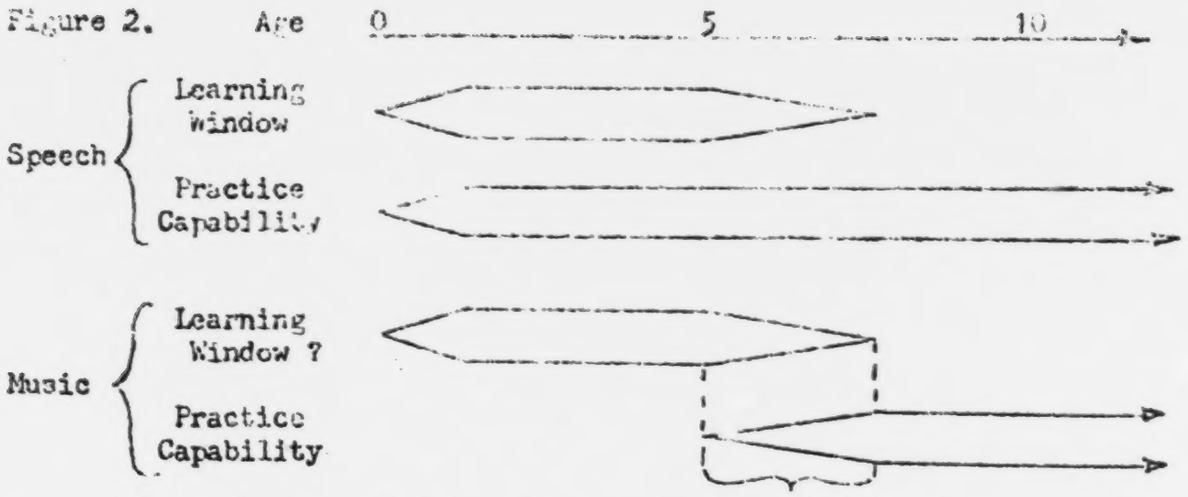
This period (to approximately age 6) may provide a "learning window" for auditory pattern classification that is closed when the hemispheric asymmetry development is completed. In the case of language, participation in language production undoubtedly provides a necessary step in the formation of a complete language generating and decoding system. Sussman (1972) states "the

versatility of the human speech production system and an increasng body of evidence suggests that speech is controlled by an intricate closed-loop feedback system. To bring about feedback control of the speech musculature, the higher neural centers should be kept constantly aware of (a) the spatial position, (b) the direction of movement, and (c) the rate of movement of the articulators. This review describes the feedback mechanisms existing within the tongue that can mediate such dynamic space-time information."

Traditional means of musical production depend upon control systems (finger, hand, and arm muscles; breath control contrary to that needed for speech) which are not sufficiently developed in the young child to allow his exploitation of musical patterns, to the degree that he explores speech. The "learning window" idea provides testable hypotheses concerning musical "talent" (musical facility demonstrated in production). There may, for example, be a critical overlap between the "learning window" and the means for exploring, and extemporizing musical patterns. The anecdotal immaturity of many professional musicians may have a close relationship to their musical ability: a delayed "closing" of the learning window may cause sufficient overlap with the developing muscle control to provide the necessary feedback contingencies (see Figure 2).

This model appears to have broad implications for many areas of human intellectual development. The role of musical pattern recognition in this model is dependent upon its possible inclusion under speech or speech-like processes. If the experiments described in this report show conclusively that some non-speech sounds are decoded in the left-hemisphere (and if, in addition they are found to produce categorical decisions: Liberman, et al., 1962) then it is unlikely that left hemisphere processing is completely limited by a finite store of feature extractors possessing some invariant relationship to a linguistic category. It is more likely that a general classification cue, such as a critical rate of change in the spectral envelope, determines the processing site. A synthesized set of sounds with variations of temporal and spectral attack parameters is being prepared in an attempt to isolate a processing cue which is not exclusively linguistic.

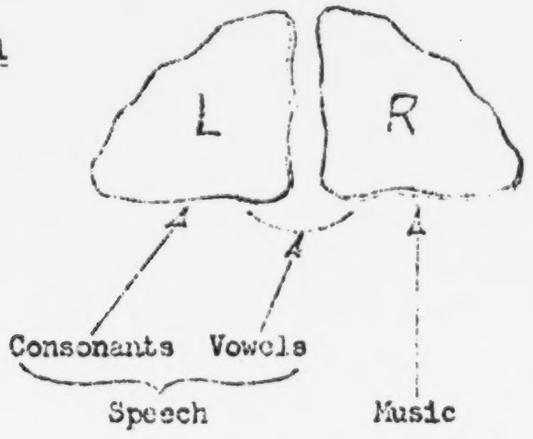
Figure 2.



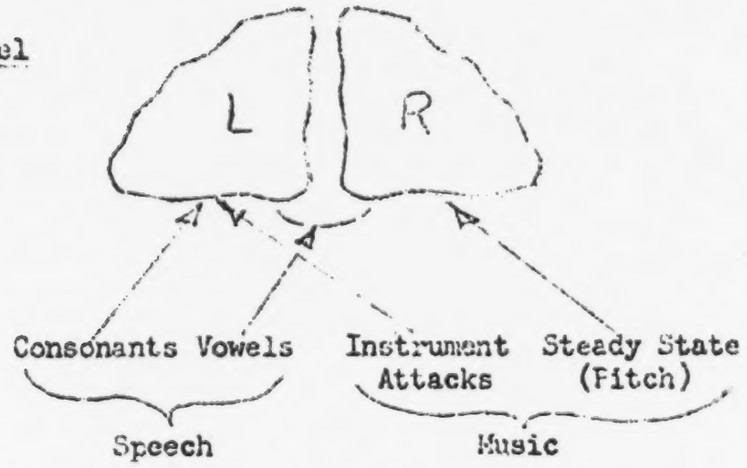
Hypothesis: A critical overlap for organizing auditory pattern and nuance structure for music?

Figure 3.

Previous Model



Proposed Model



In view of our findings taken in conjunction with those of Kimura (1967) and others, the music/speech analogy may have implications for reading problems, such as congenital dyslexia, where anecdotal accounts of unusual musical ability (non-notational) have up until now been considered as primarily compensatory adaptations. The problems of autistic children appear to be closely connected with an inability to encode language ("Autism: A Deficiency in Context-Dependent Processes?", Pribram, 1970). Anecdotes of musical precosity in some autistic children, and the role music plays in some treatment centers such as Benhaven in New Haven, Conn., indicate a close, complex relationship between speech and music for this disability. It seems to be of basic importance, however, to investigate the scope of left hemisphere auditory processing, particularly when it appears to extend beyond the range of stimuli considered in previous investigations. (See Figure 3.)

D. Systems for Performance Analysis and Modification

1. The Computer Analysis System

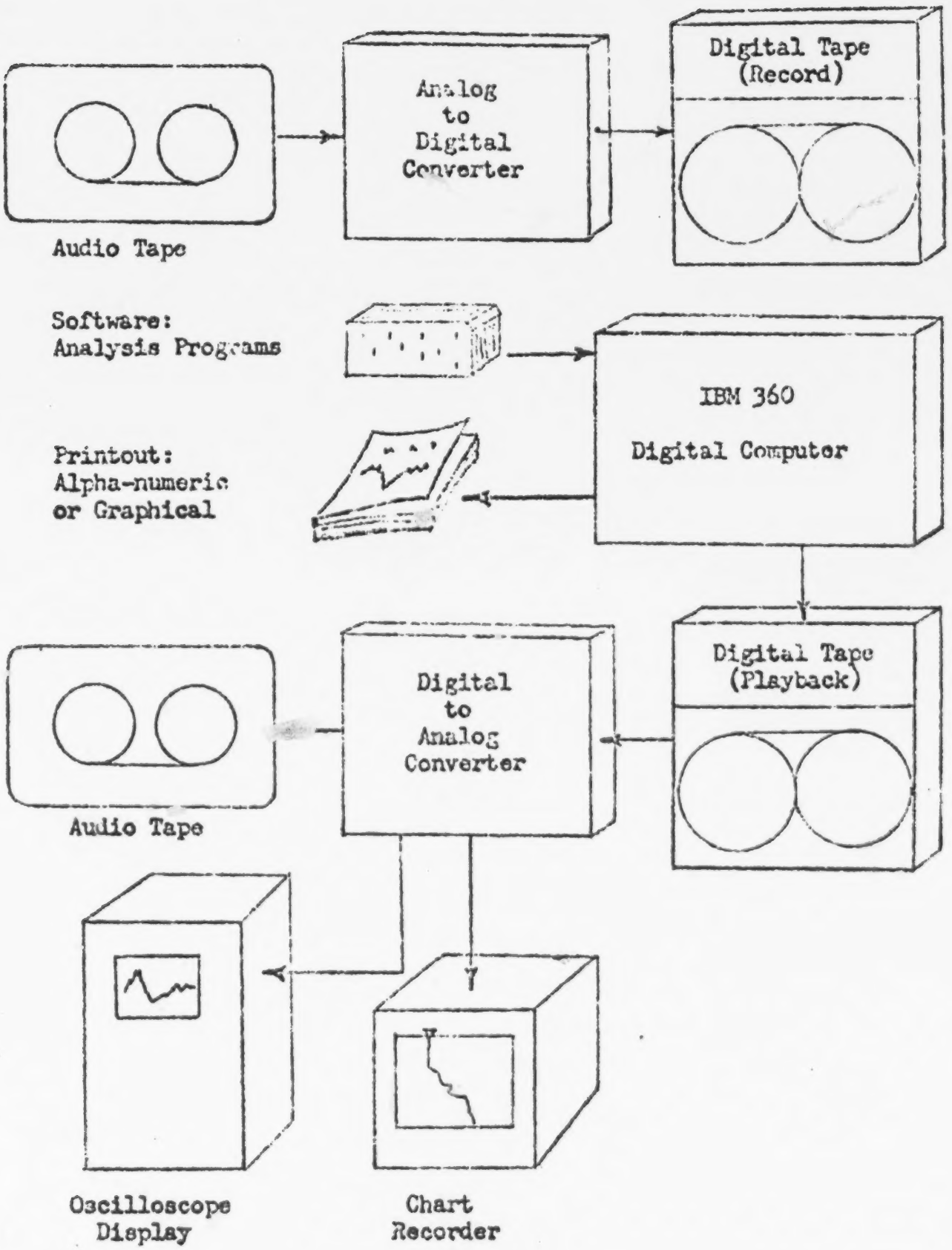
A large capacity digital computer provides a wide range of processing capabilities and is a necessity for reducing large amounts of data to useful descriptors. In order to prepare auditory material for digital processing, an analog to digital converter must be used. The analog (continuously varying) signal from an audio tape is converted to digital (discrete) values and is stored on magnetic tape in a form suitable to the input requirements of an IBM 360 digital computer. Two input formats are available:

a. Five bits + sign, at 20K words per second, which provides a 30 db dynamic range up to 10K Hz.

b. Nine bits + sign, at 10K words per second, which provides a 54 db dynamic range up to 5K Hz.

Duration, frequency, intensity, harmonic and

Figure 4.



COMPUTER ANALYSIS SYSTEM

noise analysis can readily be accomplished with the appropriate computer programs. In addition, programs for performance modification open many possibilities for investigation, since performances on traditional instruments can be altered slightly for controlled effects without destroying the naturalness of the processed sound. Figure 4 shows the system schematically, and Appendix III describes the soft ware developed for the system.

2. Tone Line Writer

The tone line writer produces a visual record on graph paper of the fundamental frequency and amplitude contours of a recorded performance, over a three octave range. (133 Hz to 1067 Hz; approximately C below Middle C to the soprano high C). A modified FM Subcarrier Discriminator (a standard telemetry device) serves as an accurate frequency meter. An octave switch, designed specifically for this system, automatically sets the proper range, substantially reducing the problems associated with tones containing strong second and third harmonics.

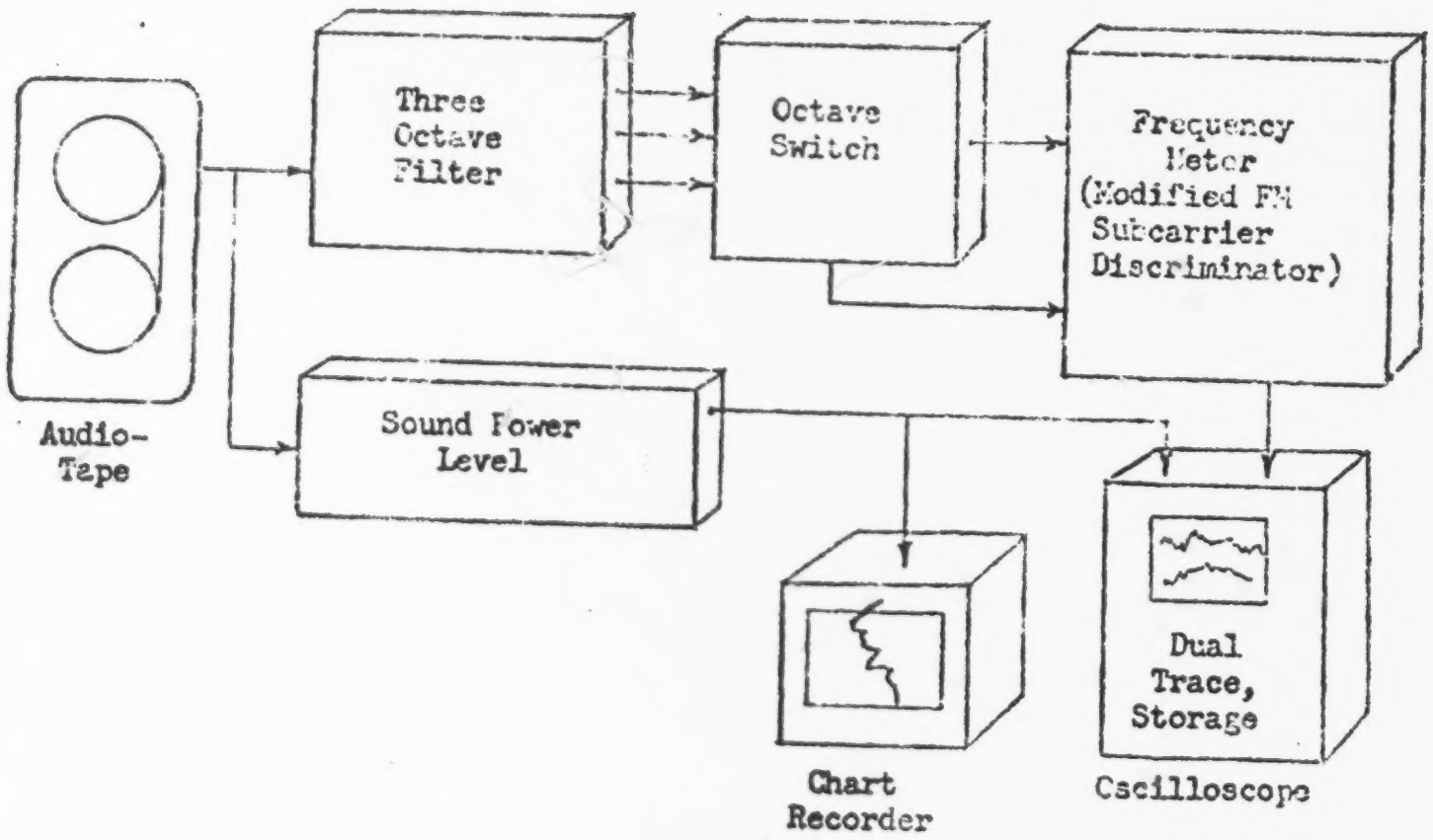
The graphs produced by this system allow visual inspection and measurement of the durations of tones and silences, and the vibrational correlates of vibrato (pitch modulation), tremelo (loudness modulation) and other vibrational characteristics. Figure 5 is a schematic of the tone line writer.

3. Tone Line Reader

The tone line reader accepts hand drawn charts representing pitch and loudness contours, and converts them to an auditory output by means of a voltage controlled synthesizer, such as the Moog. The optical scanner or "chart reader" is a modified document transmitter built by Graphic Transmission Systems, Inc. A two level signal from the scanner indicates the presence or absence of a line on the chart. The sample and hold unit determines the voltage appropriate for a line-indicating-pulse at each point in the scan, and then presents that voltage, when there is a pulse, to the proper sound control unit.

Chart paper is fed to the scanner at one inch per second, and sixty scans are made each second.

Figure 5.



TONE LINE WRITER

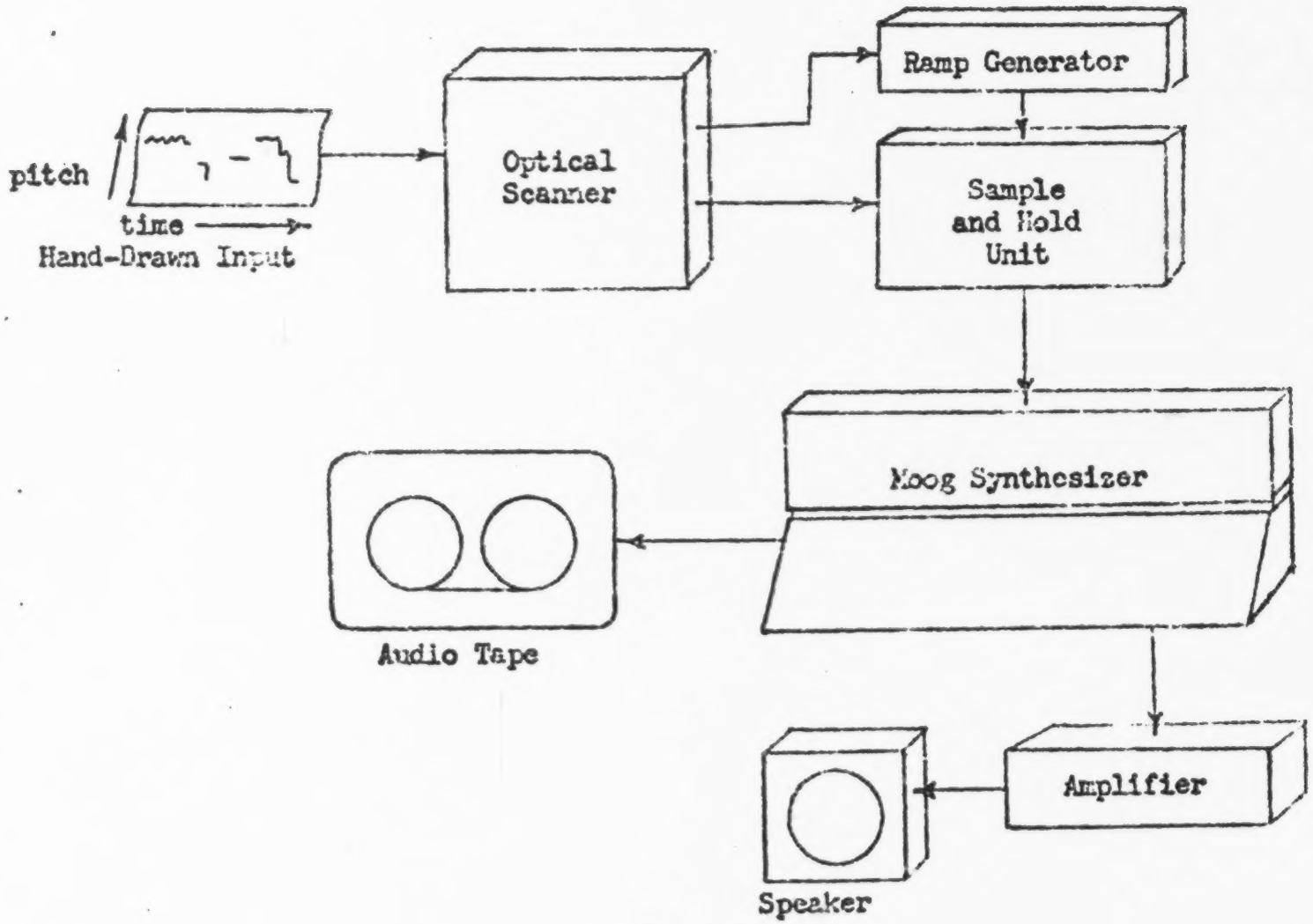
Both pitch and loudness indications can be scaled at will within the limits of the 8 1/2 inch scan width. The system is diagramed in Figure 6. The electronic interface, which converts the output of the optical scanner to a suitable control voltage, is described in Appendix IV.

4. Harmonic Synthesizer

An harmonic synthesizer with both phase and harmonic amplitude control would provide an invaluable tool for investigating the timbre changes necessary and sufficient for a particular listener response to a musical sequence of tones. Basic research on the pitch analysing processes used by the ear, the appearance of subjective tones, and the tonal characteristics of instruments would be greatly facilitated by such a device.

In view of this, a graduate student in the University of Connecticut Electrical Engineering Department was engaged to provide a prototype design for the device, and to simulate the output of the device using a digital computer. The results of this investigation are presented in Appendix V. The cost of building the synthesizer was outside the limits set by the present funding.

Figure 6.



TONE LINE READER

REFERENCES

Benade, A. (1960) Horns, Strings and Harmony. Garden City: Doubleday.

Berger, Kenneth W. (1964) Some Factors in the Recognition of Timbre. The Journal of the Acoustical Society of America. 36, 1888-91.

Bremermann, H. J. (1963) Optimisation through evolution and recombination. In Self-Organizing Systems, ed. Yovits. Washington, D.C.: Spartan Bks. p. 93.

Broadbent, D.E. & Gregory, M. (1964) Accuracy of recognition for speech presented to the right and left ears. Q.J. exp. Psychol. 16, 359-60.

Carlsen, J. (1963) Programmed learning in melodic dictation. J. of Research in Music Education, 12 (Summer), 139-148.

Cohn, R. (1971) Differential Cerebral processing of noise and verbal stimuli. Science, 172, 599-601.

Cooley, W. and Lohnes, P. (1962) Multivariate procedures for the behavioral sciences. New York: Wiley.

Deihl, N. and Radocy, R. (1969) Computer assisted instruction: Potential for instrumental music education. Council for Research in Music Education, Bulletin No. 15, Winter.

Findlayson, D.S. (1951) The Reliability of the marking of essays. British Journal of Educational Psychology, 21, 216-234.

Fletcher, Harvey. Loudness, Pitch and the Timbre of Musical Tones and Their Relation to the Intensity, the Frequency, and the Overtone Structure. The Journal of the Acoustical Society of America, 6 (1934), 59-69.

Fourier, J. (1882) Theorie Analytique de la Chaleur.

Gazzaniga, M.S. & Sperry, R.W. (1967) Language after section of the cerebral commissures. Brain, 90, 131-148.

Harris, J.D. (1952) Pitch Discrimination. J. Acoustical Society of America, 24, 750-755.

Hays, W.L. (1963) Statistics for Psychologists. New York: Holt, Rinehart and Winston.

Heller, J. (1969) Electronic graphs of musical performance: a pilot study in perception and learning. J. Research in Mus. Educ. 17, No. 2, 202-216.

Helmholtz, H.L.F. (1877) On the Sensations of Tone. Reissued: New York: Dover, 1954.

Kelley, T.L. (1947) Fundamental Statistics. Cambridge: Harvard University Press.

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychology, 15, 166-171.

---------- (1964) Left-right differences in the perception of melodies. Q.J. of exp. Psychology, 16, 355-360.

---------- (1967) Functional asymmetry of the brain in dichotic listening. Cortex, 3, 163-178.

Knox, C. & Kimura, D. (1970) Cerebral processing of nonverbal sounds in boys and girls. Neuropsychologia, 8, 227-237.

Kuhn, W.E. and Allvin, R.A. (1967) Computer-assisted teaching: a new approach to research in music. J. of Research in Music Education, 15 (Winter), 305-315.

Langacker, R.W. (1968) Language and Its Structure. New York: Harcourt, Brace & World, Inc.

Lewis, Don, and Cowan, Milton (1937) Pitch Variations arising from Certain Types of Frequency Modulation. JASA. 9, 79.

Liberman, Cooper, Harris & MacNeilage (1962) A motor theory of speech perception. In, Proc. of the speech communication seminar. Stockholm: Royal Institute of Technology.

Liberman, Cooper, Shankwieler, & Studdert-Kennedy (1967) Perception of the speech code. Psych. Review, 74, 431-461.

Licklider, J.C.R. (1956) Auditory Frequency Analysis. In: Cherry, C., Ed. Information Theory. London: Butterworth, 253-268.

Lieberman, Philip (1967) Intonation, Perception, and Language. Cambridge: MIT Press.

Lundin, Robert W. (1967) An Objective Psychology of Music. 2nd ed. New York: Ronald Press Co.

Luria, A.R. (1970) Traumatic Apasia. The Hague: Mouton Press.

Mountcastle, V.B. (1962) Interhemispheric Relations and Cerebral Dominance. Baltimore: Johns Hopkins Press.

Neilson, James (1970) Timbre and Color. The Instrumentalist, April, pp. 39-41.

Nilsson, N.J. (1965) Learning Machines. New York: McGraw Hill.

Nolle, A.W., and Boner, C.P. (1941) The Initial Transients of Organ Pipes. The Journal of the Acoustical Society of America, 13, 149-55.

Page, E.B. (1966) The imminence of ... grading essays by computer. Phi Delta Kappan, January, 238-243.

Page, E.B. and Paulus, D.H. (1968) The Analysis of Essays by Computer. Final Report, HEW Project No. 6-1318, Storrs: The University of Connecticut.

Phillips, G.E. (1948) The marking of children's essays. Forum of Education, 1, 19-29.

Pribham, K.H. (1970) Autism: A deficiency in context-dependent process? In, Proceed. of 1970 Conf. of Nat. Soc. for Autistic Children. HEW.

Rayliegh, Lord. (1877) The Theory of Sound. Re-issued: New York: Dover, 1937.

Richardson, E.G. (1964) The Transient Tones of Wind Instruments. The Journal of the Acoustical Society of America, 26, 960-62.

Risset, J. & Mathews, M.V. (1969) Analysis of Musical-instrument tones. Physics Today, 22, 23-30.

Robinson, D.W. and Dadson, R.S. (1956) A re-determination of the equal-loudness relations for pure tones. Brit. J. Apple. Physics, 7, 166-181.

Rozeboom, W.W. (1966) Foundations of the Theory of Prediction. Homewood, Ill.: The Dorsey Press.

Saldanha, E.L., and Corso, John F. (1964) Timbre Cues and the Identification of Musical Instruments. The Journal of the Acoustical Society of America, 36, 2021-26.

Saunders, F.A. (1946) Analyses of the Tones of a Few Wind Instruments. The Journal of the Acoustical Society of America, 18, 395-401.

Seashore, C.E. (1902) A voice tonoscope. Iowa State Psychology, III, 1-17.

Seashore, C.E. (1938) Psychology of Music. New York: McGraw Hill.

Sebestyen, G.S. (1962) Decision-Making Processes in Pattern Recognition. New York: Macmillan.

Shankweiler, D. (1966) Effects of temporal-lobe damage on perception of dichotically presented melodies. J. Comp. Physiol. Psychol. 62, 115-119.

Shankweiler, D. & Studdert-Kennedy, M. (1967) Identification of consonants and vowels presented to left and right ears. Q. J. of exp. Psychol. 19, 59-63.

Slawson, A.W. (1968) Vowel Quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency. JASA, 43, 87-101.

Sperry, R.W. (1964) The great cerebral commissure. Scientific American, 212, #1, 42-52 (January).

Spohn, C. (1963) An exploration in the use of recorded teaching to develop aural comprehension in college music classes. J. of Research in Music Education, 11, (Fall) 91-98.

Strong, William, and Clark, Melville (1967) A Synthesis of Wind-Instrument Tones. JASA, 41, 39-52.

---------- (1967) Perturbations of synthetic orchestral wind-instrument tones. JASA, 41, 277-285.

Stumpf, C. Die Sprachlaute, Berlin (1926).

Sussman, H.M. (1972) What the tongue tells the brain. Psychological Bulletin, 77, 262-272.

Taylor, C.A. (1965) The Physics of Musical Sounds. New York: Elesvier.

Tenney, J.C. (1965) Gravesaner Blatter, 26, 106.

Warren, R.M. & Obusek, C. (1971) Speech perception and phonemic restorations. Perception and Psychophysics, 9, 358-363.

Winer, B.J. (1962) Statistical Principles in Experimental Design. New York: McGraw-Hill.

Young, T. (1784) The Principal Phenomena of Sounds.

Appendix I

COMPUTER ANALYSIS OF MUSICAL PERFORMANCE

Warren C. Campbell

PURPOSE AND PROBLEM

This investigation is an attempt to add to the body of knowledge related to the following questions:

1. Are there operationally definably (i.e., objective, measurable) variables for which musical performance limits can be established?
2. If so, what are they and where are the limits of acceptability for various performance situations?
3. Can a knowledge of such variables be made useful in solving the problems of Music Education?

Given a set of scores assigned by a group of competent human judges to a set of student musical performances, are there objective features of the recorded performance which can be used, with suitable computer analysis, to predict the averaged judges' score?

PROCEDURES

Figure 1 provides an overview of the procedures. Two paths can be traced through the diagram, one for the human judging, the other for the computer simulation. Note that the simulation requires a representative set of the judges' scores in order to predict the judges' responses on other samples from the same population.

As shown in the computer branch of the diagram, aspects of the recorded performances were transferred to paper tape, using the equipment described in the previous chapter. The frequency meter produced a chart of performance frequency versus time, and the envelope follower presented the output voltage envelope (signal amplitude versus time).

Hand processing of the paper charts, described in this section, was used to extract the feature values for each note. Normalization and conversion of the chart values, and the subsequent reduction of the feature values to a small set of predictors were accomplished using special programs on an IBM 360 computer.

Multiple regression analysis, using these predictors

and based on a set of judges' scores produced a prediction equation which was then cross-validated on subsets of the sample. The cross validation, accomplished by correlating the predicted scores with the actual judges' scores for the subset is shown schematically at the bottom of Figure 1.

PERFORMANCE SELECTION AND PREPARATION

Fifty audio tapes, consisting of vocal and instrumental performances given at the 1968 Connecticut All-State Auditions, were auditioned to find a suitable data base for the present investigation. Of these, it was found that only the female vocal performances had a pitch range that could be processed satisfactorily using the available equipment.

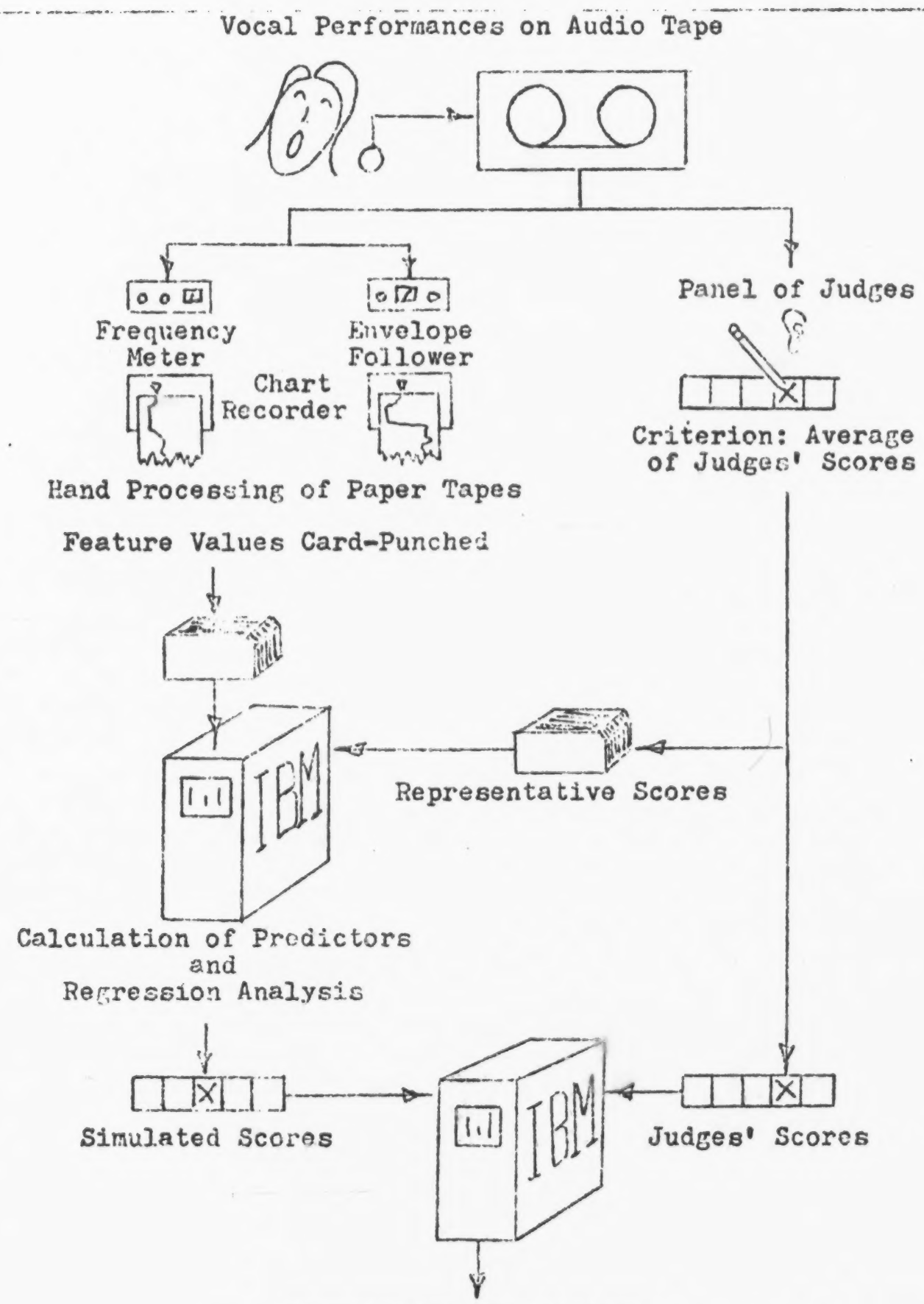
A preliminary study was conducted to determine if a stable criterion could be established for this kind of short vocal performance, and to test the judging categories, instructions and format of the judging form. Reliabilities for eleven judges in this preliminary setting ranged from .70 to .89, and were encouraging enough to warrant the present investigation.

With the information available from this preliminary study, a set of sixty-two vocal performances by sopranos and altos were taken from the original audition tapes, and were used to prepare an analysis tape and a judging tape.

Performers auditioning were allowed to choose either the aria "If With All Your Hearts" from Mendelssohn's "Elijah", or "He Shall Feed His Flock" from Handel's "The Messiah". The portions of these selections used for performance judging are shown in Appendix A. Thirteen performers sang the Mendelssohn aria. These auditions were divided, so that each phrase starting with "If With All..." was treated as a separate performance. These twenty-six performances were presented on the judging tape so that the judges could not readily connect the two segments sung by one person. The remaining thirty-six performers sang the Handel aria, bringing the total number of performances to sixty-two. The duration of the Mendelssohn performances ranged from seven to twenty-two seconds. The duration of the Handel segment ranged from thirteen to forty-two seconds.

The judging tape was made by copying the original taped performance at its nominal recording speed, preceded

Figure 1.
Overview of Procedures



by a performance number, and followed by a ten second silence. The running time of the judging tape was approximately 45 minutes. Seven copies of the original judging tape were made, so that each judge could audition the tape at his own rate and convenience.

The analysis tape was made by copying the original tapes on a variable speed recorder (Sony, Model TC-5600). By this means, the frequency range of the original performance was shifted to coincide with the range of the FM Subcarrier Discriminator.

CRITERION SCORES

Selection of Judges

The judges chosen for the panel were required to be experienced music teachers, some at the high school and some at the college level. Nine people fitting this description were contacted. Of the nine, seven were able to complete the judging task; of the other two, one was too busy, and the other found the task too difficult to do to his satisfaction. All the judges found the assignment to be an arduous one, because of the narrow range of abilities represented in the performance sample.

Directions to the Judges

The performance judges were asked for scores on a five point scale for intonation, vibrato, rhythm, dynamics, and an "overall" rating. The average of the first four categories was also calculated and used as an additional criterion value. The instructions given to the judges and the judging form are reproduced in Appendix A. The judging form contains five columns, one for each category. Each column was subdivided with headings "A" through "E", with "A" representing the best performances. The judge had only to check the letter category for each performance under each of five columns. Numerical results were card punched from these sheets with A=1, B=2, etc. Judges who completed the task were paid a fee of twenty dollars.

Each judge played the tape through at least twice. Several judges commented that the dynamics and rhythm categories were the most difficult to grade, and suggested that a diagnostic grading scheme might make their job easier.

Calculation of Criterion Scores

The grades assigned by the judges are essentially ordinal. When numerical values are substituted for the letter grades, and then averaged to provide a criterion score, an

interval scale is implied. The tacit assumption is made that the "distance" between the "A" and "B" categories is the same as the "distance" between all other adjacent categories, and that, for example, an average score of 1.5 is "halfway" between "A" and "B".

In addition, there is an assumption about the assignment of categories by the judge when the composite score used is the arithmetic mean: the scores must be considered as measurements which differ in a random fashion from a hypothetical "true" score, which represents the performance. It is only under this assumption that the average is a meaningful estimate of the performance score. If, for example, two schools of thought were represented in the panel of judges, with systematic differences between them, no single number would be a reasonable representation of the judges' opinion.

Stability of the Criterion Scores

The stability of the criterion scores can be estimated from the interjudge correlations, if the assumptions made in the previous section hold. The inter-judge correlations and a reliability estimate were calculated for the seven faculty judges. In addition, the correlation of each judges' scores with the average of the other six judges' scores was calculated as an indication of his agreement with the group opinion.

In order to test the reliability estimate, a class of nine graduate music students was asked to audition the same audio tape heard by the seven faculty judges. The average of the graduate students' scores was compared to the average of the faculty scores by calculating a correlation for each of the grading categories.

Three of the faculty judges consented to regrade the sample on a "one pass" basis, so that self-correlations could be compared with the inter-judge and inter-group correlations.

Considerable inter-category correlation was expected, based on the results obtained by Page and Paulus (1968). The stability of the less clearly defined categories was expected to benefit from this "halo" effect.

PREDICTOR VARIABLES

Selection of Features

The computer simulation of a criterion score is a problem in pattern recognition. The features of the pattern

upon which recognition is to be based are crucial to the success of a simulation. The selection, in this case, of the physical variables to be used as features depends on an assumption of a correlational relationship between the physical variables and the subjective responses of the judges.

The characteristic features chosen are listed in Table 1. These six values were calculated from measurements taken on each tone of the performances. It was expected that, while considerable intercorrelation would occur, the pitch features would be the primary intonation predictors, the power level would be the primary dynamics predictor, and that the duration features would be the primary rhythm feature. Because of the tedious hand processing necessary, no features were extracted specifically for vibrato.

Table 1

Characteristic Features, Calculated for Each Tone

1.	Initial Pitch	(cents)
2.	Middle Pitch	(cents)
3.	Final Pitch	(cents)
4.	Sound Power Level	(decibels)
5.	Duration of Tone	(seconds)
6.	Duration of Break or Glissando between Tones	(seconds)

Feature Extraction

Delays in equipment expected for machine processing made it necessary to employ hand processing techniques to extract the feature values from the recordings.

The paper tapes produced by the frequency meter and the rectifier were first segmented to define the onset and release of each tone sung. Tonal and non-tonal segments could then be measured along the time axis. The length of these segments represented the duration of a particular note sung by the performer, and the break or glissando leading to the next note. Because of noise and the decay characteristics of the system, no distinction was made between glissandi and breaks. Machine segmenting, while not a trivial problem, will undoubtably provide more consistent results with much greater flexibility.

Figures 2 and 3 show excerpts from two performances as represented on the paper tapes, with vertical lines indicating the segmenting.

After segmenting, average pitch lines were drawn for each tonal segment by estimating the center of the vibrato envelope. Three ordinate values were tabulated from each tonal segment, at the beginning, middle, and end of the average pitch line. The maximum amplitude value for each tonal segment was also tabulated. These values are also indicated in the diagrams. Data relating directly to vibrato characteristics was not used. Because of the large variations in vibrato within single performances, this data was left for machine reduction at a later time.

The tabulation of these data resulted in an N by 6 data matrix for each of the sixty-two performances, where N is the number of tones scored for each performance. The initial frequency of the original taped performance was used to calculate the change in tape speed, and to establish the original frequency of all the performance tones, since they would be affected in the same proportion by the slowing or speeding of the audio tape. The tabulated data matrix for each performance was transferred to punched cards, which were then proof-read to assure their accuracy.

Data Normalization

Since the speed of the original recorded performance had been altered to fit the frequency window of the conversion equipment, the original frequencies and durations were recovered from the data matrix using calibration factors determined during the processing of the auditory signal. The original frequencies were then normalized to a cents scale, referenced to the starting pitch chosen by the singer. This made it possible to compare pitch deviations on a scale which is independent of starting frequency. A pitch change of one cent is defined as the change from f_o to f when $f=f_o \times 2^{1/1200}$. This frequency change represents a pitch change of 1/100 of a semi-tone of the tempered scale. Visual estimation of chart lines led to confidence limits of approximately $\pm$ 4 cents on each reading, a limitation which will be eliminated when machine readout is available. The sound level chart readings (output volts) were converted to the decibel scale, referenced to the minimum signal produced by the singer. Tonal and non-tonal durations were converted to proportions of the performance duration, so that differences in tempo would not affect the predictors, but only relative differences in tone durations. These normalizations did not, of course, provide any data reduction,

but only served to facilitate the next step, which was to compare each of the performances to a standard performance.

Data Reduction and Calculation of Predictors

The approach taken for data reduction was to establish a set of standard values for each tone of the musical selection. Deviations from the standard were then calculated for each performance, and the average deviations and mean square deviations from the standard were determined for each of the six features. These twelve numbers were the basic predictor set used in the multiple regression equation. A thirteenth predictor, the ratio of total tonal duration to total non-tonal duration, was added when it was found to have low correlation with the most effective of the other twelve predictors.

The data reduction program was run twice, using two different sets of standard values. The first set was from an operational standard: the best performance in each selection category (Mendelssohn 1, Mendelssohn 2, and Handel). The second set of standard values was obtained by using a literal transformation of the notation of the musical score: the three pitch values for each tone were identical, having the value, in cents, appropriate to the score indication for that tone; the duration values were proportional to the note durations in the printed score; non-tonal durations were zero, except where rests were indicated.

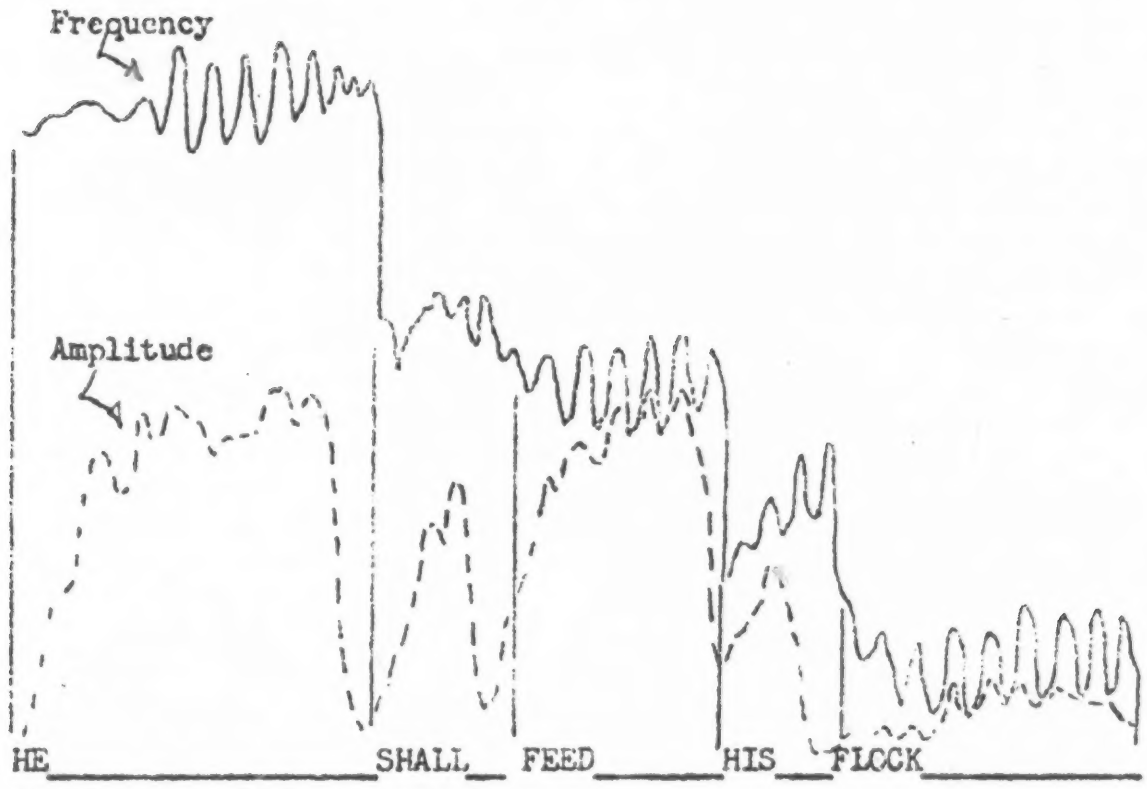
Since no score indications are given, sound power level was established by rule for the literal standard. Appropriate sound power levels appear to be, to a first approximation, pitch dependent. Therefore, decibel levels for the literal standard were made proportional to the score value for pitch referenced to the lowest tone in the sequence.

The primary differences between the operational and the literal standards can be listed as follows:

1. Pitch - The initial and final pitches in the operational standard vary considerably from the nominal pitch of the literal standard, especially on transition tones.
2. Duration - Tonal durations deviate from the nominal values, and short breaks appear between almost all tones of the operational standard. The duration ratio, since it is not referenced to a standard, is the same for both sets of predictors.

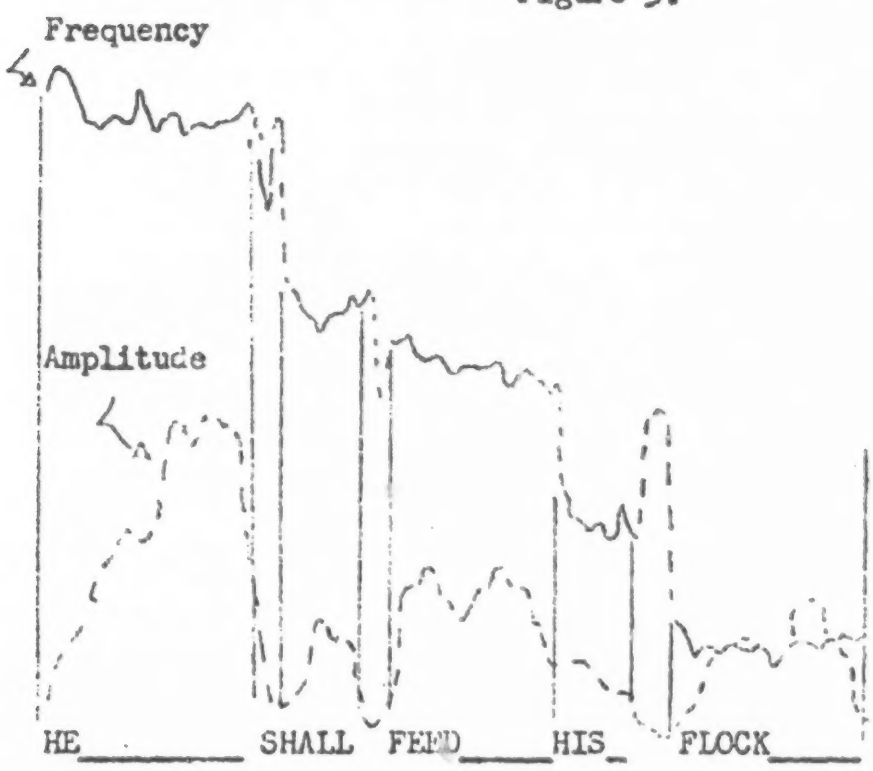
The programs used for data normalization and reduction are listed in Appendix B.

Figure 2.



Sample Chart Output: Soprano

Figure 3.



Sample Chart Output: Alto

ANALYSIS

Multiple Regression

A step-wise multiple regression analysis was run for both sets of predictors with the average of the seven faculty judges scores as the criterion. Each set of predictors was tested on the complete set of 62 performances, as well as on two partitions of the total set: a random split of 31 performances each, designated "31A" and "31B", and the natural subsets delineated by the Handel and Mendelssohn selections, designated respectively "36" and "26" in the tables, indicating the number of performances in each subset.

The use of all thirteen predictors in the linear regression equation for any subset increases the error variance accounted for, and therefore decreases the generalizability of the resulting predictor weights. In order to maintain a more nearly constant performance-predictor ratio, a selection of six predictors was made, and a new multiple regression equation using only these six predictors was calculated for each subset. The predictors were chosen by ranking, for each judging category, the predictors in the order in which they contributed to the reduction of the sum of squares. The six finally used were selected from those of the original thirteen which appeared most often in the upper half of the ranking for each of the six judging categories. The predictors used for the subsets are listed in Table 2.

There is no guarantee that these are the best combinations of predictors to select, since any one of them may be reducing primarily error variance. In that case, there would be little generalizability to other samples. However, if a set of predictors and b-weights, generated from one subset of the sample, satisfactorily predicts the scores from another subset, then these predictors may be expected to work on another sample from the same population.

Table 2

Predictors Used For Subset Calculations

Operational Standard	Literal Standard
1. Duration Ratio	1. Duration Ratio
2. Tonal Duration	2. Non-tonal Duration
3. Non-tonal Duration	3. Non-tonal Duration, Sq.
4. Non-tonal Duration, Sq.	4. Middle Pitch
5. Middle Pitch	5. Final Pitch
6. Final Pitch, Sq.	6. Sound Power, Sq.

and (say) ten outputs (for 10 bit resolution); and it requires 400 states. Clock pulse rate is 400 (100)=40KHZ. For f_1 =100HZ. And it becomes 400 (3000)=1.2MHZ. For f_1 =3000HZ.; well within the capabilities of modern logic components.

However, the device as postulated has one serious drawback - each of the 20 sequential machines has 400 states. The f_2 machine uses its 400 states to store samples of two cycles, the f_3 machine stores three cycles, etc. In order to make the device more economical, we would like to have each machine store only one cycle, resulting in the number of states getting small for the higher frequency machines. In fact, if we can manage this kind of trick, the number of states for Mi will shrink quite rapidly as i increases. Using the typical numbers above, consider that M1 stores one cycle consisting of 400 samples. Let M2 store one cycle of 200 samples (200 state machine). If we pulse both these machines at the same rate we will get two cycles out of M2 in the time it takes to get one cycle from M1, which is precisely what we want. Similarly, M4 has 100 samples, M8 has 50, M16 has 25, etc. That is, the number of states for machine Mi is 400/i. Naturally, the difficulty arises for those values of i which do not divide evenly into 400. In fact, if we wanted to design the device such that each harmonic were exactly generated by this scheme, the number of states of M1 would have to be the product of all the prime numbers in the interval (1,20), and the pulse rates required would be unthinkable.

A solution to this dilemma currently under investigation is to approximate desired behavior in the following way. Machine Mi stores the integer part of (400/i) samples. Clearly if we stopped here we would be in serious trouble, since inexact harmonics would be slightly higher in frequency than they should be, leading to successive interference effects from those pseudo-harmonics. However, the machines can be designed to inhibit the correct number of clock pulses so that at the start of each cycle of the fundamental all harmonics are in proper relation to each other. To make this clearer, let's consider M3 . We store int(400/3)=133 samples in this machine. At the end of one cycle of the fundamental, the f_3 waveform would be 1/400 ahead of where it should be. If, however, the 201st pulse is inhibited, then after 200 pulses f_3 will be (1/800) ahead, after 201 pulses it will be (1/800) behind, and after 400 pulses it will be coincident with the ideal f_3 . It is hoped that the amount of harmonic distortion introduced by such approximations will be negligible. If not negligible for a psychoacoustical laboratory instrument (in which case we can resort to equal-length machines), it might still be more than adequate for music and/or speech. Further, if necessary, the sampling theorem might be confoundable to some extent by low pass filtering the output mix at different frequencies as a function of f_1 . For example, for $f_1 < 500$ HZ. We have $f_{20} < 10{,}000$ HZ. And if we filter at 10,000 HZ. We only require 200 samples/cycle. For $500 < f_1 < 3000$, filtering at 20 KHZ, we would have a worst-case requirement of $500T_n > 40{,}000$ or $T_n > 80$. Thus with such a scheme, 200 samples would be adequate.

To reduce the number of states of the component machines, it may be possible to filter the outputs of the individual stages individually at

different cutoff frequencies, before mixing. Thus the number of states required by the sampling theorem may thereby be reduced. Of course, for an instrument of laboratory quality, much attention will have to be paid to the phase effects of such an approach.

Fig. 5 presents a set of waveforms generated directly via D/A conversion by computer simulation. For this case we have 160 samples for f_1 and are generating 11 harmonics. The "break" (pulse inhibition) positions for this set of photographs are arbitrary. In Fig. 6 we see the effects of the approximations on the filtered waveforms. Of course f_8 is a perfect sinusoid while f_7 and f_9 are not. The perceptual correlates of these distortions are under investigation. In particular, it will be interesting to discover whether apparent smoothness of a waveform effects perceived audio distortion. That is to say (Fig. 7) breaks which occur at points of zero derivative result in a waveform which appears to the naked eye to be a pure sinusoid, while breaks elsewhere are readily visible. In terms of Fourier spectrum the harmonic distortion is the same in both cases, but the visual difference is striking enough to make one wonder. In any event, we can apparently reduce the harmonic distortion as much as we are willing to pay for, and it seems that the attainment of an instrument of laboratory quality would be a relatively straightforward matter if adequate funding were available.

Figure 5. Data Samples for f_1 through f_{11}.

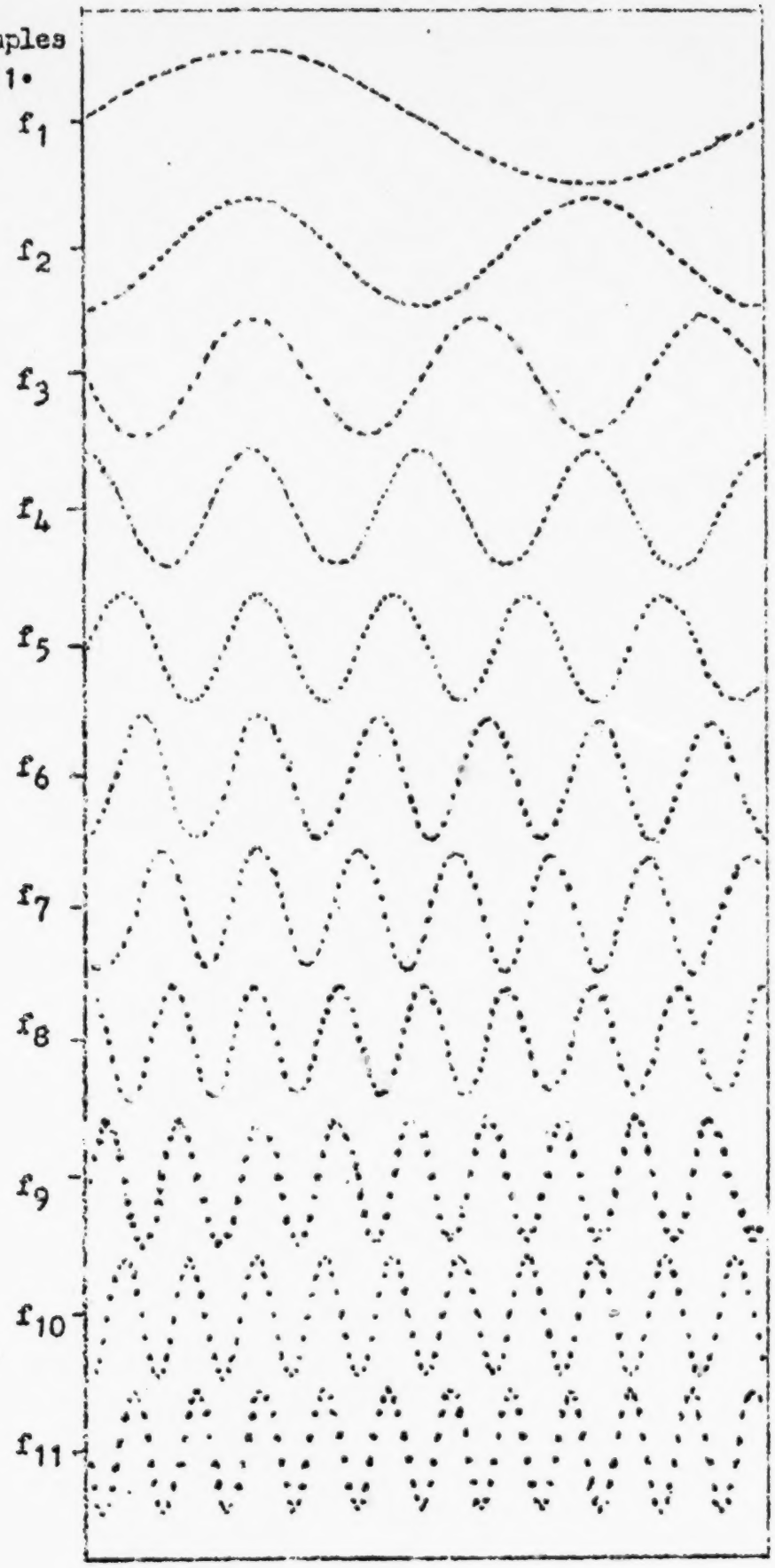


Figure 6.
Data Samples,
Before and After
Filtering

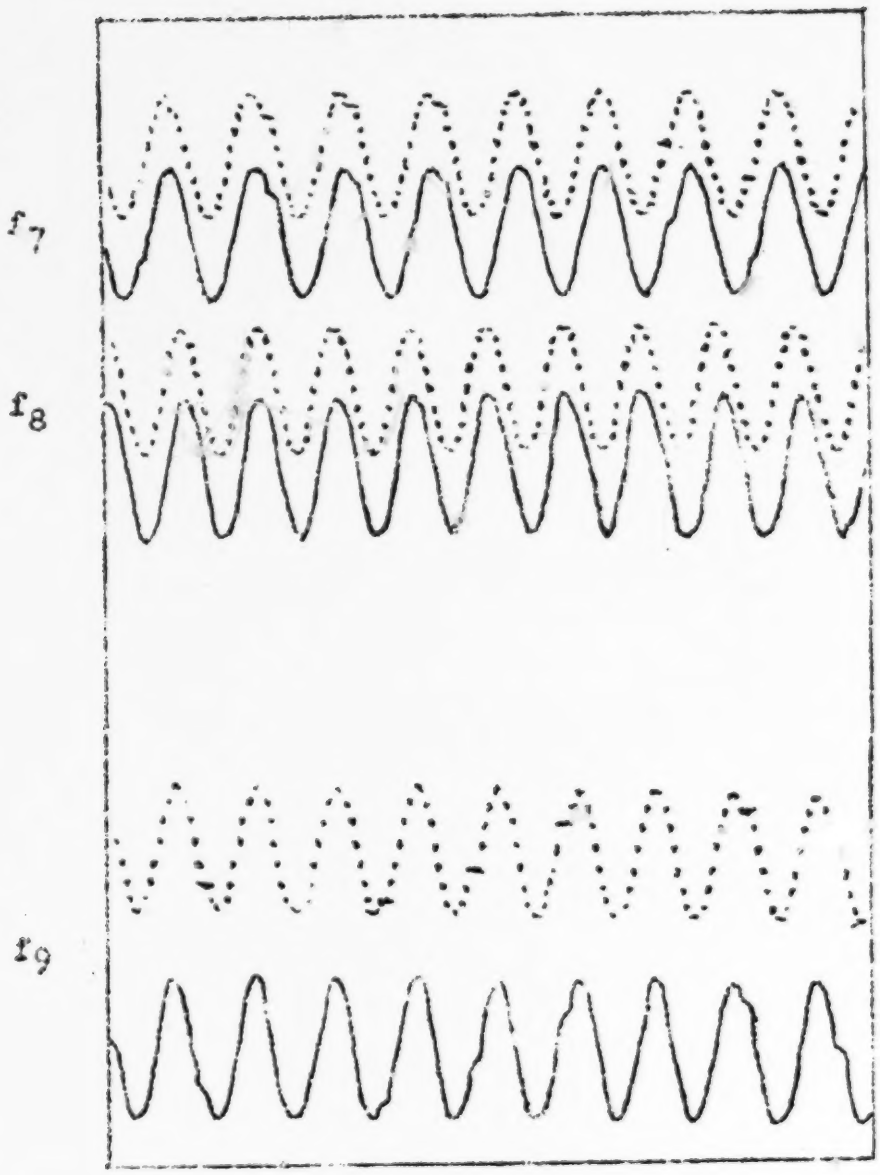


Figure 7.
Distortion at Breaks